University College London

Decoding oscillatory representations of visual stimuli in episodic memory and working memory

Anna Jafarpour

Thesis submitted for the degree of Doctor of Philosophy

2014

Declaration

I, Anna Jafarpour, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:

Anna Jafarpour

Date:

Abstract

Theories inspired by electrophysiological studies in animals suggest that the replay of past experiences plays an important role in episodic memory as well as working memory; yet very little is known about the neural characteristics of such replay in the human brain. This thesis consists of neuroimaging experiments for studying the temporal characteristics of the replay in the human brain and analytical methods for decoding replay. To that end, oscillatory neural activity patterns were recorded from healthy young adults via a non-invasive electrophysiological technique (Magnetoencephalography, MEG).

Firstly, a pipeline for decoding MEG data using machine learning algorithms was developed and proposed. Then using an associative recognition experiment, we marked the neural signature for categorical visual information (about faces and scenes) during encoding. These markers of encoding-related experiences were then used for detecting the replay during retrieval - triggered by an associative memory cue. As a result, replay was detected at about 500 ms from onset of the cue. The results suggest that episodic recollection and replay are accomplished within 500 ms.

Next, in a working memory experiment, I used item specific visual information for tracking the replay of oscillatory activity while maintaining that information. Three visual stimuli with presumably distinct cortical representations were selected (types: a face, a banana, and a chair) and presented in a sequential order. Event-related responses during encoding showed a main effect of item type and working memory load at 400-500 ms from onset of the stimuli. Using a decoding approach, it was possible to categorize oscillatory patterns related to each of the three stimulus types. These decoders are now used as markers of item specific replay in working memory during the maintenance phase. This analysis is ongoing.

Finally, we proposed a pipeline for detecting an optimal feature space for decoding MEG data at a group level because the previous pipeline relied on different features across subjects for decoding. Here the Canonical Variates Analysis of beamformer reconstructed MEG data in source space was used. Canonical Variates Analysis estimated the dependency of the selected features of MEG data to the experimental conditions and enabled multivariate decoding of MEG signal in the source space. Thus this proposed method was an optimal way for group level inference of MEG multivariate analysis.

Overall, the MEG based decoding of the representation of visual stimuli was shown in source and sensor spaces. Also, our results revealed the temporal characteristic of replay in an episodic memory experiment.

Contents

List of Figures

List of Tables

Acknowledgements

I am most indebted to my supervisor Professor Emrah Duzel for his expert guidance, support and encouragement throughout my studies. His invaluable scientific insight and his Sufi-like attitude to life have been very inspirational for me, and they are the most treasured souvenirs for me from this PhD. I am also grateful to my secondary supervisor Will Penny for his guidance and helpful consultations. I am also thankful to my collaborators, Lluis Fuentamilla and Aidan Horner, for conceptual discussions as well as for closely helping me designing and running experiments. And I also acknowledge my encouraging collaborator, Gareth Barnes, for introducing me to a new perspective toward MEG data.

Furthermore, I would like to thank other past and present members of the Queen Square: specifically Nico Bunzeck for introducing me to the SPM toolbox, Vladimir Litvak and Guillaume Flandin for sharing their insights into SPM and MEG data analyzing with me, Marc Guitart-Masip for showing me how to become more efficient in research, and Dharshan Kumaran for multiple discussions about episodic memory concepts.

I am also grateful to the director of the Institute of Cognitive Neuroscience, Prof. Geraint Rees, and the UCL graduate school for their support. I am also appreciative of the admin and technical teams' care: particularly Rosalyn Lawrence at ICN and Karen Miller, Iris Mann, and Nicole Boehnke in the University of Magdeburg. Likewise, the radiology team of the Wellcome Trust Centre for Neuroimaging was a great help in running experiments; specifically I am thankful to David Bradbury for teaching and helping me with MEG acquisitions.

I would like to also thank many others, too many to mention, who helped me

along the way, including: Eva Bauch, Pradheep Shanmugalingam, Sundeep Teki, Nahid Zokaei, Bahador Bahrami, Vincent Walsh, Markus Bauer, Hauke Hillebrandt, Rumana Chowdhary, Raphael Koster, Zeb Kurth-Nelson, Hartmut Schuetze, Dan Bush, Raphael Kaplan, James Bisby, Maria Chait, Sabine Joseph, Sylvia Vitello, Magdalena Sliwinska, and the PALS postgraduate peer group.

Finally, I am thankful to my dear aunt, Afsaneh, and her lovely family who welcomed me in London and supported me with their endless hospitality. And above all, I am filled with gratitude to my parents and my sister, Nina, for their constant love and support and for encouraging me to pursue my academic dreams.

I am grateful to my wonderful mother and father and I dedicate this work to them.

Contributions

The work in this thesis is entirely my own unless otherwise indicated. My supervisor, Professor Emrah Duzel, guided me on the studies presented in chapter 2, 3, and 4. My second supervisor, Will Penny, and my collaborator, Gareth Barnes, guided me on the study presented in chapter 5. Both of my supervisors, overseen all of my work. The episodic memory experiment in chapter 3 was designed by Lluis Fuentemilla and Emrah Duzel. Lluis Fuentemilla acquired the MEG data with my assistance. The working memory experiment was designed by me and I acquired the data. I analysed the data for the study in chapter 2 in collaboration with Aidan Horner. I also analysed the data in chapter 3 and 4. For the study in chapter 5 I adopted a combination of methods scripted in SPM12 toolbox by Gareth Barnes and Will Penny. And I also helped with developing the relevant scripts in SPM12.

Publications

- Jafarpour A., Fuentemilla L., Horner A.J., Penny W.D., Duzel E. (2014) Replay of very early encoding representations during recollection, *Journal of Neuroscience* 34(1): 242-248;

- Jafarpour A., Barnes G., Fuentemilla L., Duzel E., Penny W.D. (2013) Population level inference for multivariate MEG analysis. *PLOS One* DOI: 10.1371/journal.pone.0071305

- Jafarpour A.*, Horner A.J.*, Fuentemilla L., Penny W.D., Duzel E. (2013) Decoding the temporal dynamics of memory replay, *Neuropsychologia* 51(4): 772780 * joint first authorship

Invited talks

- 7th Annual MEG U.K. Meeting (Jan. 2013, Cambridge, UK) "Population level inference on dimensionality for multivariate MEG studies"

- 7th Annual MEG U.K. Meeting (Jan. 2013, Cambridge, UK) "Replay of representational snapshots during recollection",

- Cognition and Brain Plasticity Group, University of Barcelona (Jun. 2012, Barcelona, Spain) Studying reinstatement of encoded information at retrieval: an MEG based decoding study,

- the Oxford centre for Human Brain Activity, University of Oxford (Mar. 2012, Oxford, UK) Investigating mechanisms of working memory maintenance and episodic memory encoding using MEG based decoding,

Conference abstracts/posters

- Society for Neuroscience Annual Meeting (Nov. 2013, San Diego, USA), Jafarpour, A., Barnes, G., Penny, W., and Duzel, E. "Decoding content of working memory",

- The British Neuroscience Association's Biennial Meeting, (Apr. 2013, London, UK), Jafarpour, A., Barnes, G., Penny, W., Fuentemilla, L., and Duzel, E. "Population level inference for multivariate MEG analysis",

- Society for Neuroscience Annual Meeting (Sep. 2012, New Orleans, USA), Jafarpour, A., Barnes, G., Penny, W., Fuentemilla, L., and Duzel, E. "Categorizing faces and scenes in source space on the basis of oscillations",

- 18th International conference on biomagnetism-Biomag 2012 (Aug. 2012, Paris, France), Jafarpour, A., Barnes, G., Penny, W., Fuentemilla, L., and Duzel, E. Group-level inference of multivariate analysis in source space",

- UK MEG Annual Meeting (Jan. 2012, London, UK), Jafarpour, A., Fuentemilla, L., Horner, A.J., Penny, W. and Duzel, E. Correlations of early and later feature representations during encoding with source memory: an MEG based decoding study,

Chapter 1

General introduction

1.1 Neural representation of visual stimuli

The neural representation of visual stimuli involves multiple brain regions. Here I review aspects of the neural representation of visual stimuli in three brain regions that are relevant to my thesis - namely: the visual cortex, the medial temporal lobe, and the prefrontal cortex. These brain regions are also involved in visual episodic memory and working memory.

1.1.1 Representations of visual stimuli in the visual cortex

The occipital part of the human cerebral cortex processes mainly visual information. The visual system consists of sub-regions which are hierarchically structured (suggested by electrophysiological studies in animals (e.g. Van Essen & Maunsell, 1983) and functional MRI [1] (e.g. Grill-Spector & Malach, 2004)). The representation of visual stimuli in the visual cortex generally has three levels of complexity: low-level,

[1] Magnetic resonance imaging (MRI) is a neuroimaging technique to measure changes (mainly) in blood flow in a biological system. Functional MRI (fMRI) from the brain is assumed to reflect engagement of brain regions which function in an experiment setting.

intermediate-level and high-level (Kandel, Schwartz, & Jessell, 2000). Low-level representations include orientation, colour, contrast, disparity and movement direction. The intermediate-level refers to contour integration, surface properties, shape discrimination, surface depth, surface segmentation and object motion. High-level representations pertain to object identity. The neural networks in the early visual areas (V1, V2, and V3) are structured in a way to represent the retinotopical low-level information about visual stimuli and cover smaller receptive fields (M. Sereno et al., 1995). The visual pathway from the occipital lobe then divides into the dorsal and ventral streams. Mainly the intermediate- and high-level information of visual stimuli is characterized through these streams and they represent larger visual fields (i.e. larger receptive field).

The dorsal stream, also known as the "where/how" pathway, is located in posterior parietal cortex (area V5). It is involved in representing motion / location related aspects of stimuli, the binding of multimodal spatial representations of visual stimuli (Andersen, Snyder, Bradley, & Xing, 1997), and egocentric information about stimuli (Colby, 1998; Snyder, 2000). In other words, the motion related features of the stimuli and the location in 3D space are represented in the dorsal stream (A. Sereno & Lehky, 2011). And the ventral stream, also known as "what" pathway, is located in the inferior temporal cortex (IT/V4). This pathway represents objective properties of visual stimuli (Ungerleider & Haxby, 1994; Haxby et al., 1991), such as allocentric and categorical (what-ness) information. For example, a part of the fusiform gyrus - in the temporal lobe - activates more strongly for faces than other visual categories. This area, hence, predominantly represents faces and thus is called the fusiform face area (FFA) (Kanwisher, Stanley, & Harris, 1999). Similarly, the parahippocampal place area (PPA) is activated more for places than objects; and the extrastriate body area (EBA) represents body parts more than objects (Haxby et al., 1991; Peelen &

Downing, 2005). It is notable that although high-level representations seem to reflect abstract complex categories, single cell recordings from monkeys suggest that rather abstract representations still carry the low-level features of visual stimuli, such as relative positions of stimuli in the visual field (Kravitz, Kriegeskorte, & Baker, 2010; A. Sereno & Lehky, 2011).

The categorical representations in IT sub-regions are not absolute. Representations of visual stimuli are not limited to the region which maximally responds, but they are distributed (and overlapping) over the ventral visual pathway (Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999). As such, FFA is not only more active for faces relative to objects, but also it is more active for a non-living doll face than an arm (living body part). For that reason, one should study the representation of visual stimuli considering a network that is distributed over a larger brain area rather than focussing on a single area.

In order to investigate which type of visual categories have the most distinct distributed neural representations, Kriegeskorte et al. (2008) applied a dissimilarity matrix of neural representations. They compared the similarity of neural activity for various categories of visual items in the human IT (using fMRI). In other words, the dissimilarity matrix was used to find out what features of visual stimuli generate more distinct (spatial) distributed neural representations. It emerged that the two main clusters of visual stimuli are animate and inanimate. Then in a hierarchical fashion, within animate group clusters of human and non-human features, and within both human and non-human clusters body and face clusters were identified. Also within inanimate cluster they found natural and artificial visual stimuli (Kriegeskorte et al., 2008).

1.1.2 Representations of visual stimuli in the medial temporal lobe

The medial temporal lobe (MTL) of the human brain consists of the hippocampus, perirhinal, entorhinal, and parahippocampal cortices. The hippocampus itself is a collection of subfields: CA1, CA3, Subicular complex and dentate gyrus (Amaral, 1999; Squire, Stark, & Clark, 2004). In 1971, neurons in CA1 were discovered to represent the positions of freely navigating rats in their environment (i.e. "place cells") (O'Keefe & Dostrovsky, 1971). This significant discovery indicated the existence of sparse representations of the environment in the hippocampus. Later, Kreiman, Koch, and Fried (2000a) recorded from multiple regions including MTL of the human brain while the participants perceived different categories of visual stimuli. They showed that the firing rate of hippocampus neurons has category-specific patterns and is particularly strong for spatial configurations (Kreiman et al., 2000a). Interestingly, the representations of category-specific visual stimuli in the MTL are not limited to the physical perception of a visual stimulus but rather, similar representations also emerge when subjects imagine visual items (i.e. visual imagery) (Kreiman, Koch, & Fried, 2000b).

Representations of environments in the hippocampus are multimodal. Sperling et al. (2003) showed representations of faces and names binding in the anterior regions of the hippocampus and observed a role for the hippocampus in representing associative information (Sperling et al., 2003). Such multimodal representations are known as memory traces of the event (Marr, 1971). The formation of the memory trace is critical as it enables the later retrieval of experienced events (Marr, 1971).

The hippocampus is a unique structure which facilitates memory formation and retrieval. It is rich with recurrent connections within its sub-regions and its repre-

sentation of events is sparse (Waydo, Kraskov, Quian Quiroga, Fried, & Koch, 2006). These two attributes provide particular characteristics to hippocampal representations. The most significant characteristics are a low overlap between representations and a high capacity for memorizing with fast learning rate (Treves & Rolls, 1994). Therefore, the hippocampus can save representations about a large number of events as engrams (memory trace) (Lashley, 1950; Bruce, 2001). In fact the representations of events in the hippocampus are shown to be involved in multiple cognitive processes including episodic memory (Sutherland & McNaughton, 2000; Moscovitch et al., 2005) and working memory (Moscovitch et al., 2005).

The neural structure of the hippocampus enables two important information processes on memory representations of event: pattern completion and pattern separation. When a partial cue activates a subset of neurons in a neural assembly, which represents the cued event, consecutively the recurrent connections reactivate the whole neural assembly. This process is called pattern completion. On the other hand, pattern separation benefits more from the sparseness of event representations. The sparseness of representations leads to precise and low-overlapping neural patterns, which enables emergence of separate representations for similar events, i.e. pattern separation (Yassa & Stark, 2011).

1.1.3 Representations of visual stimuli in the prefrontal cortex

The prefrontal cortex has been a key area implicated in the process of actively keeping information in mind (PFC; section 1.3). The PFC is thought to be able to act like an information buffer for representing selected relevant information (to be later used by executive functions, for example; Rainer, Asaad, & Miller, 1998; Miller &

Cohen, 2001). Along this line, recording from primates showed that the PFC represents the quantity of visual items (Nieder, Freedman, & Miller, 2002) as well as the categorical information of visual stimuli (e.g. PFC represents cats and dogs information when the task is to classify a probe to a cat or a dog; Freedman, Riesenhuber, Poggio, & Miller, 2001). The representations in the PFC are, however, unlike the IT cortex (Nystrom et al., 2000). Possibly representations of events in the PFC are distributed, dynamic (e.g. Baeg et al., 2003), and indicative of task demands (Haynes et al., 2007) rather than items. Studying representations of tasks / items in monkey-PFC supports this notion (Averbeck, Crowe, Chafee, & Georgopoulos, 2003; Everling, Tinsley, Gaffan, & Duncan, 2006).

1.2 Episodic memory

Episodic memory is a form of long-term memory. One proposal for classifying human long-term memory is into implicit and explicit memory (Schacter & Graf, 1986). Implicit memory refers to procedures (e.g. motor skills and habits, such as how to drive), conditioning, and priming. Implicit memory affects the behaviour without being consciously or intentionally retrieved (Graf & Schacter, 1985). Explicit memory, on the other hand, refers to specific facts (i.e. semantic memory) and events (episodic memory) that are consciously retrieved.

Episodic memory is then a type of explicit memory for autobiographical events (Tulving & Donaldson, 1972) which allows the conscious retrieval of contextual details of past experiences (Tulving, 1985). Encoding is the first stage of episodic memory formation, when the external event is perceived and provokes an internal representation. The representation then goes through the process of consolidation

which renders its neural representation (trace) long-lasting. In this process, the representation is compared to the existing knowledge already stored. This knowledge can then be updated and new links between old and new experiences can be established (Morris et al., 2006; Morris, 2006; S. Wang & Morris, 2010; Nadel, Hupbach, Gomez, & K, 2012). Thus, the consolidation process is closely linked to the process of retrieval, as it brings up the existing knowledge about present events. Retrieving episodic memory itself engages representations of event contents; for example, where and when the event happened (Nadel & Hardt, 2011; S. Polyn, Natu, Cohen, & Norman, 2005).

1.2.1 Testing episodic memory recognition

Experimental designs for studying episodic memory usually consist of two phases: encoding and retrieval. The encoding phase is when stimuli are presented for certain duration of time. The depth of encoding can be instructed or manipulated. For instance, emotion can manipulate the depth of encoding (Cahill & McGaugh, 1998), or instructions such as attending to a certain aspect of stimuli (e.g. the font of a word; F. I. Craik & Lockhart, 1972) or attending to stimuli in relation to subjects' previous experience (e.g. semantic meaning of a word F. I. Craik & Lockhart, 1972) affects the depth of encoding. Of course encoding is not independent from retrieval processes because encoding engages representations of previously experienced events (F. Craik, 2002).

At the retrieval phase, memories are usually tested via form of free recall, cued recall, or recognition (Tulving, 1985). In free recall, subjects are asked to recall experienced events freely (without a cue; e.g. naming the items in a list which was presented during encoding). In cued recall, subjects are asked to recall items with

specific attributes (e.g. naming items in the list which started with letter 'A'). And in a recognition test, previously presented (i.e. old) items or unstudied items (i.e. new) are used; subjects are asked to indicate if a probe is old or new (Rugg, Cox, Doyle, & Wells, 1995). Also at retrieval, the subjective measure of recognition can be questioned. For example, subjects are asked to rate their confidence level or to select 'remember' or 'know' in a remember/know paradigm (whether they remember the probe or they have a general feeling of knowing it) (Woodruff, Hayama, & Rugg, 2006; Voss & Paller, 2009).

Everyday experience informs us that there are two ways to recognize that a stimulus came across us on a previous occasion. In some occasions, we experience a general feeling of familiarity which indicates that the presented event has been experienced before. On other occasions, we remember or recollect contextual details from a previous encounter. Both routes, familiarity and recollection, lead to recognition (Yonelinas & Parks, 2007; Rugg & Yonelinas, 2003; Squire, Wixted, & Clark, 2007; Eichenbaum, 2008). In case of recollection, source memory tests are also designed to query memory for contextual details. In source memory tests, subjects are asked to not only report their memory of the probe but also report the associated contextual information (e.g. Davachi, Mitchell, & Wagner, 2003).

1.2.2 Episodic memory brain network

Understanding the underlying mechanisms of episodic memory requires identifying the structures which contribute to the memory formation and recognition. Using fMRI, the neural activity for successful memory formation during the encoding process of associative memories (B. Kuhl, Rissman, & Wagner, 2012) and retrieving them (Sohn, Goode, Stenger, Carter, & Anderson, 2003) are localized to lateral pre-

frontal cortex and the medial temporal lobe.

Functional MRI studies showed that specifically the hippocampus has an important role in episodic memory formation (Wagner, 1998; Davachi & Wagner, 2002; Squire et al., 2004; Saksida & Bussey, 2010; Cashdollar et al., 2009; Tubridy & Davachi, 2011). In fact, the importance of the hippocampus for episodic memory-like tasks was discovered by Scoville and Milner (2000) in 1957, through the pattern of memory impairment shown by an amnesic patient, H.M. Although H.M. had an intact memory of remote events in past, he suffered from anterograde amnesia for newly experienced events. Later, MRI scans showed that H.M. had a unilateral lesion in the medial temporal lobe of his brain (Corkin, Amaral, Gonzalez, Johnson, & Hyman, 1997). Also in line with the fMRI studies which showed engagement of PFC in the episodic memory (Donaldson, Petersen, Ollinger, & Buckner, 2001; Leshikar & Duarte, 2012; Jenkins & Ranganath, 2010), (PFC) lesions studies showed that recognition with familiarity and recollection are impaired in patients with PFC lesions and in patients with disrupted connectivity between the hippocampus and PFC (Knowlton & Squire, 1995; Aggleton & Brown, 1999; Duarte, Ranganath, & Knight, 2005).

Another brain area which contributes to episodic memory is the parietal cortex (Sohn et al., 2003; Berryhill, Phuong, Picasso, Cabeza, & Olson, 2007). The parietal cortex is generally involved in the representation of egocentric space and attention (Colby & Goldberg, 1999; Schindler & Bartels, 2013). Although patients with parietal lesions do not show severe episodic memory impairment, this area is proposed to enable an attentional focus and feedback regarding previously encoded memories rather than current perceptions (Cabeza, Ciaramelli, Olson, & Moscovitch, 2008; Ciaramelli, Grady, & Moscovitch, 2008). Furthermore, the parietal cortex is also

associated with the subjective confidence of participants in recognition (Yonelinas, Otten, Shaw, & Rugg, 2005; Henson, Rugg, Shallice, & Dolan, 2000).

As a final note, visual episodic memory system is not limited to PFC, MTL, and the parietal lobe. For example, the amygdala has been repeatedly shown to be involved in memory encoding and retrieval for arousing and emotional stimuli. This in turn affects the quality of episodic performance (Cahill & McGaugh, 1998; Dolan, Lane, Chua, & Fletcher, 2000).

1.3 Working memory

Working memory is the active maintainance and manipulation of information for a certain duration (usually for few seconds) (Baddeley, 1992, 2010). The maintenance characteristics of working memory makes it act like a memory buffer and the manipulation characteristic of working memory links the process to attention (Downing, 2000; Awh, Vogel, & Oh, 2006) and executive functions (Baddeley, 1992).

Working memory is usually tested in form of a free recall, cued recall or delay-match-to-sample experiment. Some well-known working memory tests are n-back task, in which subjects are asked to answer if the probe is the same as n-items-back in the list, the Sternberg test (Sternberg, 1966), where subjects are asked to answer 'Yes' if the probe was in the sequence and 'No' if it was not, and digit-span test, where subjects are asked to repeat the digits on a list after a short time.

1.3.1 Brain networks associated with information maintenance

Multiple brain regions are involved in working memory maintenance. Firstly, single-cell recordings from the PFC in monkeys identified neurons which represent and maintain spatial and visual information in visual working memory (Rainer et al., 1998). Similarly, functional MRI showed the engagement of the human PFC in working memory (Braver et al., 1997). The PFC activity during the maintenance period reflects the number of items held in working memory (working memory load). It also reflects the quality of retrieval processes or mnemonic representations (Rypma & M, 1999) with overlapping representations (Reynolds & O'Reilly, 2009). The degree of overlapping representations is suggested to depend on the number of items in the working memory (Bays, Catalao, & Husain, 2009).

Recently the MTL has been shown to play a role in organizing working memory. Initially it was suggested that the hippocampus is not necessary for working memory performance because H.M. had normal working memory performance for few seconds (Corkin et al., 1997). However, patients with bilateral hippocampal injury and severe deficits in episodic memory also show impairments in short-term associative recognition memory tests, even with only a few seconds interval and specifically with large memory load (Olson, Moore, Stark, & Chatterjee, 2006). Overall, it is conclusive that reinstatement of associative memories relies on the hippocampus, irrespective of whether the retention interval is short or long (Cashdollar et al., 2009; Hartley et al., 2007; Hannula & Ranganath, 2008). In addition to the hippocampus, the parahippocampal cortex is involved in working memory for novel items; and working memory for novel items is suggested to be different from familiar ones (Stern, Sherman, Kirchhoff, & Hasselmo, 2001; Hasselmo & Stern, 2006).

Working memory maintenance is not limited to the MTL and PFC. In fact, using fMRI Cohen et al. (1997) showed that the parietal cortex is also involved in working memory maintenance. And using decoding neuroimaging data, the maintenance of event-specific representations have been detected in the human visual cortex (Harrison & Tong, 2009).

1.4 Magnetoencephalography

The possibility of directly recording neural activity from the human brain using intracranial electroencephalography is very much limited to certain clinical situations, e.g. during the presurgical or intraoperative phase of epilepsy surgery in patients with therapy-resistant epilepsy - who are generally not a good representative of healthy population. Due to such constraints, we require a non-invasive neuroimaging tool for measuring healthy brain activity with high temporal resolution.

Dendritic activities generate electrical currents and simultaneously magnetic fields (Murakami & Okada, 2006). When the dendrites' activities - mostly of pyramidal cells - are rhythmic and have perpendicular orientation on the scalp, the magnetic fields are detectable by magnetoencephalography (MEG) (Murakami & Okada, 2006). Thus, magnetoencephalography is a non-invasive electrophysiological neuroimaging technique which measures the magnetic fields generated by electrical current flow in sulci due to evoked or spontaneous neural activity in the living human brain.

MEG signals are in the order of femtoteslas (ft; 10^{-15}); their temporal resolution is excellent with sampling rates usually in the order of a few milliseconds (ms; 10^{-3})

and a spatial resolution of several millimetres (mm,10^{-3}). Although MEG captures the overall activity of neural populations over the scalp, source localization of oscillatory activities into their neural assemblies is possible. Differences in sensor-sensitivity toward sources enables source localization and monitoring changes in neural assemblies over time (Dalal et al., 2009; Hoogenboom, Schoffelen, Oostenveld, Parkes, & Fries, 2006).

Oscillatory MEG-signals are characterized by their frequency, amplitude, and phase. These characteristics reflect the rate, intensity, and timing of oscillatory neural activities. Event-related MEG changes are recorded/analysed mainly in two ways: via oscillatory synchronization and desynchronization. The former refers to increased activity, mainly via excitatory connections between the neural assemblies, which leads to increases in oscillatory power relative to baseline (e.g. pre-event activity). In contrast, desynchronization of neural assemblies, mainly via inhibitory connections, is associated with decreases in oscillatory power (Lopes da Silva, 2006).

The event-related changes are usually extracted as follows. First, MEG signals are averaged over multiple trials, phase-locked to the onset of given experimental events (i.e. event-related evoked response). The same analysis is done on the signals before event onset (baseline) or in another event (or condition). The averaged signals between conditions are then compared. This method is based on the assumption that event-related activity evokes an oscillatory response in addition to on-going oscillations (which is assumed to be unrelated to the event; noise). In some cases the oscillatory activity can be induced in different phases; hence, averaging the signals, as mentioned above, diminishes these induced event-related responses. Therefore, induced responses are calculated by averaging the extracted power of oscillatory responses over trials, irrespective of the phase of the signals (Tallon-Baudry

& Bertrand, 1999), i.e. event-related induced response.

Classically, oscillations are divided into frequency bands: delta (<4 Hz) theta (4 to 8 Hz), alpha (8 to 12 Hz), beta (12 to 30 Hz), gamma (30 to 70 Hz) and high gamma (>70 Hz) (Nunez, 1981). Event-related changes of oscillatory activities can be within these frequency bands. For example, in general theta oscillation is thought to coordinate the functional cooperation of brain regions (Jones & Wilson, 2005). And specifically, oscillatory changes in theta frequency are evident in episodic memory and working memory systems (Duzel, Penny, & Burgess, 2010).

1.5 The temporal characteristics

1.5.1 Neural signature of categorical representation in the visual cortex

Electroencephalographic studies using visual stimuli have shown that visual representations are formed in the occipital cortex at about 80 to 200 ms from the onset of the stimulus. The timing of representations is dependent on the category (VanRullen & Thorpe, 2001). For example faces are shown to generate a large and fast event-related field response - at about 170 ms from the onset of the image (M170; J. Liu, Harris, & Kanwisher, 2002; Gao et al., 2013). This event-related component is source localized in the ventral stream (e.g. M170 event-related component for faces is in the fusiform gyrus; Deffke et al., 2007). Yet, representations of visual stimuli are not limited to the ventral stream (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005) and possibly have representations in other regions such as the MTL and PFC (possibly at later time points).

1.5.2 Neural signatures of episodic memory

At retrieval, multiple significant differences between event-related responses of recognition hits (correctly recognized old items) and correct rejections (new items) are identified (Duzel et al., 2003). The first old/new difference occurs at 200 and 300 ms across occipital channels. Then the N400 component emerges at 400 to 500 ms in the anterior temporal lobe and a late positive component (LPC) is evident at 500-700 ms at temporal, parietal and frontal channels (Duzel et al., 2003).

Beside event-related components, activity in theta, gamma, and (sometimes) alpha and beta frequency bands have been shown to correlate with episodic memory processes. iEEG recordings from the MTL of epilepsy patients showed that synchronizations (100-300 ms and again 500-600 ms after stimulus onset) and then desynchronization (1000-1100 ms) of gamma activities (40 Hz) between the rihinal cortex and the hippocampus predict subsequent memory performance for words (Fell et al., 2001). Accordingly, gamma activity is thought to reflect the formation of local representations of events (Fell et al., 2001).

Hippocampal activity in theta frequency is identified during episodic memory encoding and retrieval (Sederberg, Kahana, Howard, Donner, & Madsen, 2003; Osipova et al., 2006; Duzel et al., 2010). More specifically, Klimesch, Schimke, and Schwaiger (1994) acquired electroencephalography (EEG) signals from the human brain and showed that theta power is stronger in episodic memory tests than semantic memory tests. Furthermore, at episodic memory retrieval theta (4-5 Hz) power is stronger for recognized old probes than new ones at 200-500 ms (Duzel, Neufang, & Heinze, 2005), and theta synchronization during encoding predicts subsequent behavioural performance (Weiss, Muller, & Rappelsberger, 2000).

Other frequencies, like alpha and beta frequencies, are also linked to successful recognition (Klimesch, Doppelmayr, Schwaiger, Winkler, & Gruber, 2000; Duzel et al., 2003; Mormann et al., 2005; Sederberg et al., 2007; Hanslmayr, Spitzer, & Bauml, 2009) and the interactions between these oscillatory activities reflect memory performance. For example, in an iEEG study from the MTL of epilepsy patients, increased gamma power and its synchronization with theta oscillations during encoding predicted successful encoding (Fell et al., 2003). Also, at retrieval, an increase in theta oscillatory power in pario-temporal areas and occipital gamma power was associated with successful memory recognition (Osipova et al., 2006).

1.5.3 Neural signatures of working memory maintenance

Alpha (9 to 12 Hz) activity from parieto-occipital areas reflects working memory load during the maintenance period (Jensen, Gelfand, Kounios, & Lisman, 2002; Tuladhar et al., 2007). The power in theta frequency increases with memory load (Jensen & Tesche, 2002). Therefore, alpha and theta oscillations are suggested to play a role in the representation of information during maintenance.

Working memory is thought to be organized by theta (4 to 8 Hz) and gamma (30 to 80 Hz) oscillations (Lisman, 2010). A well-established model of working memory (based on the candidate physiological mechanism) predicts that theta oscillations (possibly from the hippocampus) coordinate cortically distributed information during the maintenance. This model relies on neural network oscillations and reinstatement of neural representation of events, i.e. replay (Buzsaki & Draguhn, 2004).

1.6 Models of episodic memory and working memory based on neural representations

In 1982, Mishkin (1982) proposed a mechanistic functional network for episodic memory, inspired by electrophysiological studies in monkeys, in which representations of visual stimuli have a key role. In this model representations of visual stimuli are stored in the cortex. The representations in turn can activate cortico-limbic-thalamocortical circuits. The activation and reactivation of these neural representations in the circuit then enables recognition (Mishkin, 1982).

Here, I review two prominent models of episodic memory: the temporal context model and the complementary learning system model - which are relevant to my thesis. These models are suggested based on biological attributions of the neural networks and the representations of stimuli. The temporal context model (TCM) is based on actively maintaining stimulus representations, which enables the recognition after a short delay (Howard & Kahana, 2002) and Complementary Learning System (CLS) model relies on changes in the weights of neural network connections, which enables long-term memory formation (McClelland, B, & OReilly, 1995).

1.6.1 Complementary learning system

The CLS model (McClelland et al., 1995) emphasizes the role of the hippocampus and the neocortex in shaping episodic memory. Based on the CLS model, the neocortex represents and stores a model of the world and the hippocampus rapidly and automatically stores the pattern of cortical activities so that they can be recalled later through a pattern completion process (McClelland et al., 1995; OReilly & Rudy, 2000). The slow learning rate of the cortex allows adaptation of previous representations of events by the new ones.

This model supports the dual-process theory of recognition (Yonelinas et al., 2002). It assumes that in order to keep a track of the similarities between events, the cortex stores overlapping representations of past events. Accordingly, the cortex detects a novel stimulus based on dissimilarities between cortical representations of old and new stimuli. Whereas, the hippocampus stores distinct representations of events in order to pattern-separate them; and it stores new representations without disturbing the representation of similar old events. This model suggests that at recognition, cortical representations compute a scalar familiarity signal and the hippocampus processes recollection.

In the CLS model, the entorhinal cortex (EC) acts like an interface (input/output) of cortical representations to/from the hippocampus, which stores neural representations of stimuli. In this model, the hippocampus is considered in three layers - dentate gyrus (DG), CA1 and CA3 - where CA3 stores a sparse representation of an episode and CA1 remaps the sparse representation of CA3 to EC representation of the event. This circuit enables pattern completion in the hippocampus. Due to sparse representation in CA3 and sparseness of the neural representations in DG, the hippocampus also pattern-separates events with overlapping features.

1.6.2 Temporal context model

In the temporal context model, neural representations of events include the representation of context, which drifts over time (Howard & Kahana, 2002). The representation of contextual drift enables temporal information about events in their representations. The significance of TCM is that the representation of previous events drives the contextual representation. As a result, during retrieval a memory

cue can recall a chain of events, which are linked via the similarities in their representations (like dominos of partial cues). The formulation of event representations in TCM includes (1) the information about the previous event (2) the updating of context representation based on the previous and current information (3) the current stimulus representation. Therefore, in a recursive way, TCM jumps back in time in order to generate a current representation of an event (K. Norman, Detre, & Polyn, 2008).

The idea of a contextual drift in TCM can explain some findings of episodic memory and working memory experiments. For example, in a delay-match-to-sample experiment the most recent samples are better recalled (Bjork & Whitten, 1974); according to the model, the temporal context representation embodied with the probe representation is most similar to the representation of the most recent sample; thus the probe acts like a partial cue (K. Norman et al., 2008). Similarly, in a free recall test, the events which happened in an adjacent temporal order are most probable to be recalled one after the other compared to those that happened more remotely in time (Howard & Kahana, 2002).

1.6.3 Working memory maintenance buffer

According to Cashdollar et al. (2009) episodic memory and working memory share similar mechanisms. Along that line, the above mentioned models of episodic memory can also account for aspects of working memory. Indeed, implementation of TCM needs an active maintenance buffer (like working memory). The buffer in TCM holds contextual information of episodic memory using a simple recurrent network (Howard & Kahana, 2002; S. M. Polyn, Norman, & Kahana, 2009); PFC can model such a buffer. In fact, O'Reilly and Frank (2006) have used the biological structure

of PFC to model the mechanism for maintaining information. In the model of working memory, the maintaining memory-stripes keep information (like PFC) and an information gate (similar to basal ganglia) controls the input and output (to/from the PFC) (O'Reilly & Frank, 2006). Another model, based on rodents, however, proposed that the entorhinal cortex acts like a TCM buffer through persistent neural activity (Egorov, Hamam, Fransen, Hasselmo, & Alonso, 2002). Egorov et al. (2002) specifically suggested that while PFC is important for monitoring familiar stimuli, the MTL is important for an active maintenance during delay period. Also, sustaining information in EC assists the mechanism for memory encoding and consolidation Egorov et al. (2002).

A possible physiological mechanism for the maintenance of information in working memory relies on 'replay' of information (Buzsaki & Draguhn, 2004). The central hypothesis is that during maintenance memories are replayed through gamma oscillations (distributed neocorticaly) that are phase-locked to theta oscillations (coordinated by the hippocampus). This mechanism works as a cortical multi-item buffer (Sirota et al., 2008; Jensen & Colgin, 2007). In line with the model, theta oscillations are associated with the replay of item-information during maintenance in the human brain (Mizuhara & Yamaguchi, 2011; Fuentemilla, Penny, Cashdollar, Bunzeck, & Duzel, 2010; Poch, Fuentemilla, Barnes, & Duzel, 2011). However, characteristics of the replay of neural patterns during the maintenance of sequential information are not yet understood in detail.

1.7 Thesis overview

Thus far, in animals the replay of neural activity has been functionally related to the ability to retrieve associative information in episodic-like memory, (A. Lee & Wilson, 2002; Foster & Wilson, 2006) and to the ability to maintain information online in working memory (H. Lee, Simpson, Logothetis, & Rainer, 2005; Siegel, Warden, & Miller, 2009). To test these theories in the human brain it is necessary to measure brain activity with an acquisition technique that has a high temporal resolution (e.g. with MEG) and to apply analytical techniques that allow decoding the representation of neural activity during encoding and later during retrieval (when hypothetically similar neural patterns are reinstated). The aim of my thesis is to develop the required analytical pipeline and investigate temporal characteristics of the neural replay in episodic memory and working memory.

In chapter 2, I review MEG decoding studies and propose an optimal pipeline for multivariate analyses of MEG data (using pattern classifiers). Next, I use the developed pipeline to investigate the replay of neural representation at episodic memory retrieval. In chapter 3 an associative recognition experiment is adopted for studying the timing of the replay at recollection. In chapter 4, I investigate the representation of sequential information in a working memory experiment. Later, the data set from chapter 3 is used to propose an improved decoding approach for group-level inference of multivariate analysis using MEG data (chapter 5). The final chapter contains the general discussion on the thesis. The next section reviews methods for MEG data collection and preprocessing.

1.8 Methods

1.8.1 MEG data acquisition

The MEG data were acquired using a CTF 274 channel Omega system with a sampling rate of 600 Hz. In this machine, the 274 channels are located in a helmet shaped liquid helium container (Dewar), in which the participants head is placed. Each channel is a superconducting quantum interference device (SQUID) which is a sensitive low noise detector of magnetic fields. SQUID converts the changes in magnetic field, which are picked up by the SQUID coil, into voltage allowing the detection of weak neuromagnetic signals. The SQUID operates based on superconductance and the liquid helium in the Dewar helps maintaining an operative temperature. The MEG system is located in a shielded room to minimize the interference of external magnetic disruptions and noise detection procedure is performed every time before scanning. The main environmental noise detected in this thesis is at 50 Hz, which is generated by electrical equipments. After participants are comfortably seated in the MEG, head localizer coils were attached to the nasion and 1 cm from the left and right pre-auricular (in line with the corresponding outer canthus) to monitor head movement during the recording sessions.

1.8.2 Data pre-processing

Here I used Statistical Parametric Mapping (SPM) toolbox [2] based on MATLAB analytical software for analysing the data. I took the following steps to pre-process the signals. Firstly, the data for each participant was converted to the SPM data structure format. Then it went through Notch filtering, in order to filter out the 50 Hz environmental noise (stop pass filter 49 to 51Hz). Before epoching the data

[2]SPM8 in chapter2 and chapter3 and SPM12b for chapter4 and chapter5

based on the event of interest, data were also filtered to include only the frequencies of interest, if it was necessary for the project (see the methods session of each study chapter for details). Afterwards, the data were epoched, baseline-corrected if necessary (by subtracting the mean power of the epoch signal), and labelled by experimental conditions. Then epochs in which signal intensities exceeded a threshold of 1500fT were considered to contain artifacts and were consequently rejected from further analyses.

1.8.3 Event-related responses

ERFs were extracted by comparing the averaged signal to the signal in the baseline period, e.g. before onset of the event. Afterwards, the ERFs were compared relative to the experimental conditions or memory performance. For example, at retrieval the ERFs for memory hits were compared with correct rejections (e.g. Duzel, Picton, et al., 2001; Neufang, Heinze, & Duzel, 2006) or relative to the confidence in recognition (e.g. Neufang et al., 2006). Possible differences in evoked responses were in amplitude, timing or the source of the signals.

Induced responses were extracted after transforming data into the time-frequency (TF) domain before averaging across trials/samples. The TF data was then averaged and compared between conditions (Tallon-Baudry & Bertrand, 1999). Potential differences in induced responses were considered in timing, frequency, power or the source of the effect.

For the TF transformation, I used the wavelet transformation algorithm. A five-cycle complex Morlet wavelet 1.1 was adopted to extract the time-frequency information on each single-trial data, for each participant and each channel. A complex Morlet

wavelet is formulated as

$$W(f,t) = (2\pi\sigma_t^2)^{1/2} e^{\frac{-t^2}{2\sigma_t^2}} e^{2i\pi f_0 t}, \tag{1.1}$$

where the relation $2\pi\sigma t f_0 = 5$ in here (C. Tallon-Baudry, Bertrand, Delpuech, & Permier, 1997). Accordingly, the TF representation of signal $s_k(t)$ (signal for trial k and frequency f and time t) is computed as $F_k(f,t) = w(f,t) \times s_k(t)$, where $\times$ denotes the complex convolution (C. Tallon-Baudry, Bertrand, Peronnet, & Pernier, 1998).

1.8.4 Source reconstruction

A challenge in MEG analysis is to estimate the location and orientation of neural activity in a given temporal window, i.e. inverse problem. The inverse problem is ill-posed and does not have a single solution; thus, the source reconstruction is estimated. One estimation approach is based on spatial filtering, i.e. beamforming. In my thesis (chapter 5) I used a forward model which was estimated using a single homogenous shell model of the head shape of each subject (Mosher, Leahy, & Lewis, 1999). A forward model (Nolte & Curio, 2000) estimates how much the neural activity in each brain region is detected by each channel on the scalp. I used the linearly constrained minimum variance (LCMV) beamformer to implement source localization for each trial. LCMV algorithm finds potential sources (vectors) which generated the neural activity in the frequency of interest and then compares them in an experimental design to identify the most likely source of the activity (G. R. Barnes & Hillebrand, 2003). Other examples of spatial filters are to reconstruct the source data in each condition separately using dynamic imaging of coherent sources (DICS) (Gross et al., 2001) or to assign a scalar estimate of optimal activity at each voxel

using synthetic aperture magnetometry (SAM) (Robinson & Vrba, 1999).

Chapter 2

Decoding categorical oscillatory representation of visual stimuli

2.1 Precis

Through this chapter I developed a pipeline for multivariate pattern classification (MVPC) of electromagnetic data. I decoded oscillatory representations of visual categorical information (faces and scenes) in alpha-beta-gamma frequency ranges. The performance of pattern classifiers was tested by cross-validation approach. An accurate classification algorithm was then used in the subsequent chapters [1] .

2.2 Introduction

Decoding representations of visual stimuli (based on MEG or EEG data) has been mainly used for the Brain and Computer Interactions: 'reading minds' (e.g. Besserve et al., 2007; Soto et al., 2009; Llera, van Gerven, Gómez, Jensen, & Kappen,

[1]This chapter derives in part from: "Decoding oscillatory representations and mechanisms in memory" Jafarpour A, Horner AJ, Fuentemilla L, Penny WD, Duzel E. (2013) *Neuropsychologia*; 51(4):772-80. doi: 10.1016/j.neuropsychologia.2012.04.002.

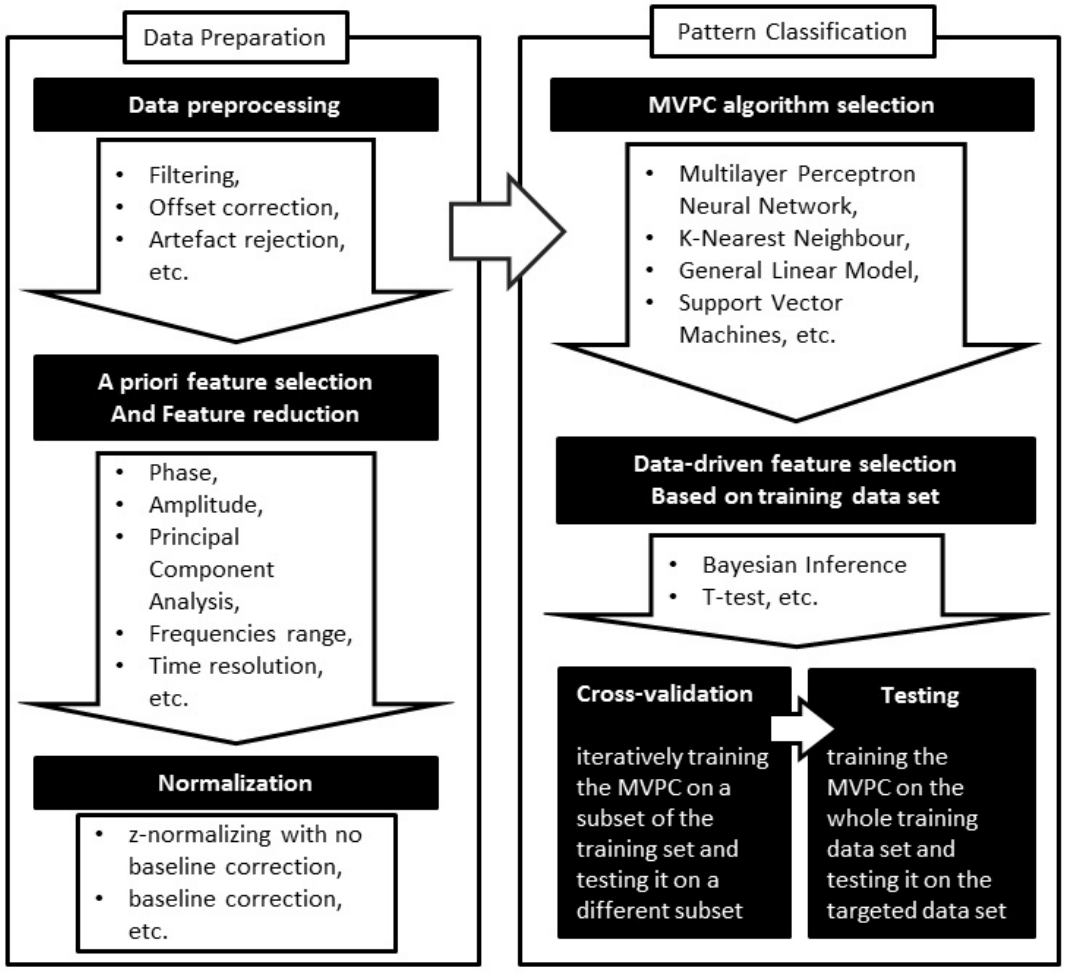

Figure 2.1: Decoding steps

2011). Recently decoding MEG data has been applied in cognitive neuroscience (e.g. W. Wang et al., 2010; Newman & Norman, 2010; Carlson, Hogendoorn, Kanai, Mesik, & Turret, 2011). For example, Fuentemilla et al. (2010) decoded the replay of visual item information during working memory maintenance in a delay-match-to-sample test.

Generally, decoding consists of two main steps: data preparation and pattern classification (Figure 2.1). Data preparation includes the preprocessing and feature reduction. And pattern classification includes selecting the algorithm, data driven feature selection and cross-validation, and finally testing the validated pattern classifier. The quality of multivariate pattern classification (MVPC) performance depends on the data pre-processing steps and the classification algorithm. In this chapter, I tested multiple methods for feature selection, baseline-correction, and pattern classification.

2.2.1 Feature selection

For each subject, the time-frequency represented MEG data has three dimensions: time point, frequencies and channels. Feature selection refers to the methods for reducing the number of features. Such methods change the feature space and/or select the relevant features, in order to facilitate robust classification. In my thesis, I was interested in representations in fine temporal windows, therefore, I mainly used frequencies and channels as features. The number of features that are potentially used can be very large (e.g. in here: with 274 channels and 40 frequency bands, there are potentially 10960 features per time point); thus, feature selection is essential.

Two main parameters can constrain the features of interest: the hypothesis parameters and the pre-processing techniques. A hypothesis driven constraint on time-

resolution is, for example, when the neural reactivations coupled with the phase of theta frequency is studied. Accordingly, the temporal resolution of the reactivated data (which is detected by decoding) should be higher than half of the duration of a theta-cycle (for a 5 Hz wavelet, this should be higher than 100 ms).

Also a pre-processing technique can constrain the sampling rate or temporal resolution of time-frequency transformation. For instance, the number of cycles in a wavelet transformation is a constraining parameter: small number of wavelet cycles leads to better temporal resolution but compromises the resolution for frequency information.

In case of decoding the differences between representations of stimuli, it may be more important to achieve a good separation of frequency features; however, as Newman and Norman (2010) discussed, the low temporal resolution (with long wavelets and low frequency bands) can lead to potential smearing of MVPC performance into the baseline period. Here, I applied the commonly used 5-cycle wavelet transformations.

Hypothesis driven feature reduction

One simple feature selection method is a conceptual apriori restriction of frequency space. For example, for studying offline reactivation of neural patterns coupled to oscillations at theta frequency, neural reinstatements must be brief. Therefore, the decoding algorithm should be restricted to fast oscillations in the beta/gamma frequency band.

For example, in the decoding approach used in (Fuentemilla et al., 2010), the goal was to test the hypothesis that reactivation of memory representations is phase-

locked to slow oscillations in the theta-frequency range. Hence, it is evident that MVPCs trained to detect representations should not include features in the theta frequency range and should be restricted to those fast frequencies (beta and gamma) that are hypothesized to be phase-coupled with theta. More generally, however, such a restriction may not be appropriate. For example, in Newman and Norman (2010) the motivation for decoding was to understand how information representation at encoding determines a specific form of priming. Thus, it was more appropriate to train MVPCs without such an a priori restriction.

A second possibility for feature selection is to restrict sensor or source space on the basis of a priori anatomical hypothesis. For example, in Chapter 5 I use fusiform gyrus and primary visual cortex as regions of interest (ROI) and decoded only the signal from ROIs. This type of feature selection can be achieved using various source modelling tools (G. Barnes, Litvak, Brookes, & Friston, 2011).

Data-driven feature selection

After feature selection based on hypotheses, data-driven feature selection steps can follow. One possibility is to use a univariate statistic at each channel and time-frequency step in order to select those features that would constitute the independent variables (i.e., the inputs) for MVPC. The statistics can then be thresholded at standard values to select relevant features.

Another feature selection procedure is Principal Component Analysis (PCA) (Bishop, 2006a). PCA is a mathematical transformation of data from its original representation into a lower dimensional linear subspace. The selected subspace explains the largest amount of variance in the data. Thus, the number of principal components

is equal or less than the number of original variables (Bishop, 2006b). For example Manning, Polyn, Baltuch, Litt, and Kahana (2011) used PCA representations of iEEG data for comparing representations and for decoding.

Importantly, the data-driven feature selection step should not involve any data from the testing set. For instance, in Fuentemilla et al. (2010), feature selection was conducted on all trials presented during encoding, but MVPCs were used to classify delay activity later on. If the goal of that study would have been to make inferences on how representations at encoding relate to subsequent memory or for validating the MVPC performance, data driven feature selection should have been performed separately in each cross-validation iteration, thus excluding the testing data set.

2.2.2 Baseline correction

Most studies normalize power in EEG and MEG data with respect to some baseline activity. One of the commonly used baselines is the activity immediately preceding stimulus-onset. This *single-trial baseline correction*, however, can potentially compromise classification accuracy because the baseline period is not devoid of information. It is likely that information related to some form of prediction or expectation of the upcoming stimulus or the continued rehearsal of the previously seen item may be present. Depending on the exact nature of the experiment, some types of baseline selection could improve or corrupt classification. For instance, in a random sequence of A and B (e.g. AABABBBAAAABABB), approximately half of the A trials are preceded by A and the other half are preceded by B. If there is rehearsal of the preceding item (A or B) in the baseline periods of A, single-trial baseline correction would confound half of the A stimuli with (baseline) representations related to B and vice versa. This could compromise classification accuracy. Therefore,

it may sometimes be more appropriate to think about alternative methods of baseline correction (e.g. taking running averages of baselines of neighbouring items of A).

One other way of baseline-correcting is relative to the condition, *condition-specific baseline correction*. In this case the average pre-stimuli onset activity for each condition is used for baseline correction (e.g. submission from the whole epoch Fuentemilla et al., 2010). The advantage of this method is that in the above mentioned example, for instance, the baseline would have average information about As and Bs. However, if the baseline contains information about the forthcoming stimuli, the condition-specific correction can generate a systematic bias for As or Bs during the baseline.

Another alternative is to avoid baseline normalization altogether. This can be done for example by z-normalizing the signal in each trial rather than baseline correction (Newman & Norman, 2010). The advantage is that any information representation at the baseline will not affect the whole epoch after correction. Furthermore, using z-normalization might enable the implementation of MVPC to EEG/MEG data recorded from long time intervals in which baseline periods are compromised, such as in sleep studies or long resting states periods. In this chapter I decode a data set using these three methods for baseline correction and compared the quality of classification.

2.2.3 Classification algorithms

In previous EEG and MEG studies of memory multilayer perceptron (MLP) (Fuentemilla et al., 2010) and regression (Newman & Norman, 2010) classifiers have been used. Various other classification algorithms can be used in principle. The classification

algorithms I used for decoding were MLP, Sparse Bayesian Classifier (SB), and Support Vector Machine (SVM).

In this chapter (I) I first examine the effect of methods of baseline correction on MLP classification performance using trial-specific and category-specific baseline correction. (II) Then I tested different methods of feature selection, namely using PCA and t-test for feature selection using MLP. (III) Finally, I used MLP, SB, and SVM for decoding.

2.3 Methods and materials

2.3.1 Empirical data

MEG data used in this chapter was from an associative recognition test where images of unfamiliar faces and scenes were shown to the subjects for 2 seconds during encoding and the images were later tested at retrieval phase (see Chapter 3 for the details of the experiment). The data from 11 subjects were used. One subject was excluded due to low behavioural performance. For each subject, there were 10 scene and 10 face images in each run of the encoding phase. There were 6 runs overall, but for 3 subjects one run was excluded for technical reasons (see Chapter 3). This data set was used to find an optimal pipeline for decoding category-specific representations.

2.3.2 Feature selection and baseline correction

Feature selection

I tested two sets of frequencies for the TF analysis. I analysed gamma/beta band frequencies which were 52 wavelets from 12 to 90 Hz (12 to 45 Hz in steps of 1 Hz and 56 to 90 Hz in step of 2 Hz.) and lower-gamma/beta/alpha band which were 38 wavelets from 8 to 45 Hz (in steps of 1 Hz). The TF-transformed data was then downsampled from 600 to 300 Hz sampling rate. Then, 17 time bins were selected to include the baseline and 1000 ms after the onset of the images. The time bins were centred at: -153, -87, -20, 47, 113, 180, 247, 313, 380, 447, 513, 580, 647, 713, 780, 847, and 913 ms relative to stimulus onset. The classifiers were then trained separately for each subject and time-point.

Baseline corrections

The TF representation was afterwards baseline corrected (baseline: -200ms to 0ms from onset) with three different procedures. (1) Category-specific baseline correction: In this procedure, the power at the baseline period was averaged across trials of the same category. Hence, a baseline value for the face category and another baseline for the scene category were calculated. Then the trials were corrected (subtracted and divided) relative to their baseline. (2) Trial-specific baseline correction: In this procedure, each trial was corrected using the average power in the baseline period in that trial. (3) Normalized: the TF data was not baseline corrected in this procedure; instead at each time, frequency and channel the power was normalized across trials. To select the most appropriate baseline correction method, I used thresholding (univariate t-tests over features) for feature reduction and MLP for classification.

Feature reduction

After selecting the appropriate baseline correction method, using MLP I tested two feature reduction methods: PCA and thresholding. PCA was performed on average power across trials in each frequency and channels and at each time point. Different numbers of principal components for the classifier were tested on data from one subject. I used 30 principal components for this analysis after all on the group of subjects. In another feature reduction procedure, univariate analysis was performed on spectral power at particular frequency, time and channel that would constitute the independent variables for the classifier. Those features which were found to be significantly different between categories by a two-tailed paired Student's t-test ($\alpha = 0.05$) were selected.

2.3.3 Cross-validation

The performance of a pattern classifier at each time point was evaluated using a cross-validation approach. In cross-validation, trials were divided into training and testing/validating groups. In each cross-validation iteration the models predicted the category of the left-out trials. The performance was calculated as the average across the cross-validation iterations. I used 10 fold cross-validation: in each cross-validation iteration the pattern classifier is trained on 90% of the trials and 10% of trials are left out for testing.

The performance of each classifier was assessed separately by comparing the group-level accuracy with 50% chance level using a one-tailed t-test. Significance thresholds were then corrected for multiple comparisons over time for the final results (see). In this chapter this was implemented using a Bonferroni correction - which is a very conservative correction (more appropriate correction - random field theory - is ex-

plained in Chapter 3).

2.3.4 Classification algorithms

The following classification algorithms were used in this chapter. I used MLP for decoding data with different feature selection methods (baseline-corrections, frequency ranges, and feature reduction). I used all the following classification algorithms for decoding z-normalized TF data in 8 to 45 Hz using thresholding feature reduction.

MLP: Multilayer Perceptron Classifier is a non-linear classifier which consists of three layers, input, output and hidden layer. The neural network used here consists of 4 hidden units and one output unit, with feed-forward logistic function connections. Each unit in hidden and output layer had a bias vector. The weights of the connections were initialized from a zero mean, unit variance isotropic Gaussian distribution. The new input was assigned to a class according to the number generated in the output unit. (This classifier has been used before for a similar data type; see Fuentemilla et al. (2010)).

SB: Sparse Bayesian Classifier is a statistical model which uses Bayesian inference for sparse classification (Tipping & Faul, 2003). For input-target pairs $\{x_n, t_n\}$, $n = 1..N$ and $t_n \in \{0, 1\}$, the SB classifier assumes a Bernoulli likelihood distribution, $p(t \mid w)$, and adopts a logistic sigmoid link function $\sigma(y) = \frac{1}{(1+e^{-y})}$ applied to $y(x) = \sum_{m=1}^{M} \{w_m \phi_m(x)\}$, where $\phi_m(x)$ is a basis function, w is a weight vector, and y is the output. This leads to the following likelihood:

$$p(t \mid w) = \prod_{n=1}^{N} \sigma\{y(x_n; w)\}^{t_n} (1 - \sigma\{y(x_n; w)\})^{1-t_n}. \tag{2.1}$$

In the training steps the mode of posterior distribution, $p(w \mid t, \alpha) \propto p(t \mid w)p(w \mid \alpha)$, is estimated using a Laplace approximation procedure for finding the most probable weights, w. Here α is the prior precision of weight w. Therefore, for a new x_* , prediction of t_* is concluded from the following distribution:

$$p(t_* \mid t) = \int p(t_* \mid w, \alpha)p(w, \alpha \mid t)\mathrm{d}w\mathrm{d}\alpha. \tag{2.2}$$

The prediction based on the maximum posterior (MP) parameter estimates is

$$p(t_* \mid t, \alpha_{MP}, \sigma^2_{MP}) = \int p(t_* \mid w)p(w \mid t, \alpha_{MP})\mathrm{d}w. \tag{2.3}$$

Here, I used the Sparse Bayesian Model toolbox developed by Mike Tipping (http://www.relevancevector.com) for linear classification of encoding images into faces and scenes. In this case the target contains the class label of zero (face) or one (scene). The maximum number of iteration for finding the most probable parameters is 1000 iterations.

SVM: Support Vector Machine with a linear Kernel (Vapnik, 2000a) is a classification algorithm which groups inputs into classes based on a linear decision boundary it finds through learning. During learning procedure, SVM finds a decision boundary similar to a linear classifier which additionally has largest classification margin (largest distance between the groups members and the boundary). Here, I used a linear kernel, which means the identical data was used as input of the classifier. The SVM algorithm is implemented in the Matlab bioinformatics Toolbox.

2.4 Results

2.4.1 Baseline correction

I applied three different baseline correction procedures, namely category-specific baseline correction, z-normalization, and trial-specific baseline correction. And I used MLP for classification. As results of cross-validation, when the data was trial-specific baseline corrected, none of the MLP classifiers performed significantly better than chance ($P > 0.05$ uncorrected for multiple comparison). Z-normalization baseline-correction resulted in above chance classification at about 200 ms, and category specific baseline corrected data was classifiable 300 to 900 ms better than chance (50%).

The baseline corrections were not done within cross-validation iterations. This above chance classification at 300 to 900 ms using category-specific baseline correction can be due to a systematic influence of categorical baseline correction. In other words, the training data set already includes information about the testing data set. Such caveat is minimally valid to z-normalization, because the normalization was done across all trials irrespective to the category of the stimuli.

2.4.2 Feature selection

The MLP classifier performance after PCA (12-90 Hz data) was not significantly better than chance; however, at 180 ms (uncorrected $P = 0.0094$) and 246 ms ($P = 0.0394$) the accuracy was above 50% (Figure 2.2). Such low performance might be due to large variation of features within categories (larger than between categories).

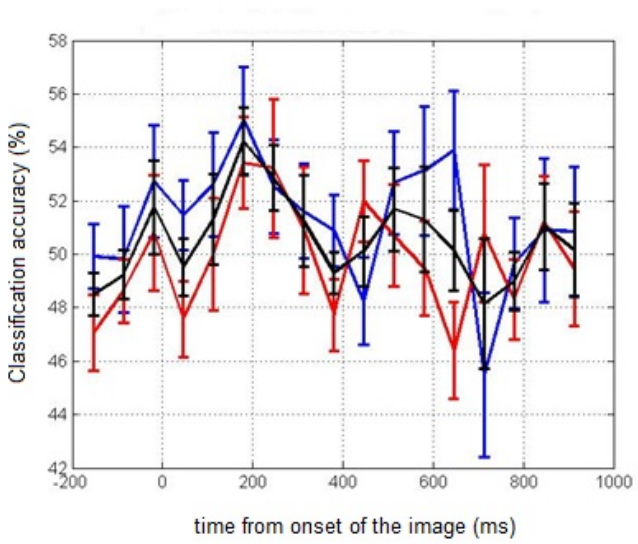

Figure 2.2: Feature reduction using PCA Accuracy of MLP classifiers on PCA data set (12-90 Hz), the error bars indicate standard error of the mean: The blue is the classification accuracy for faces, red for scenes and black for the averaged. (The Averaged classifications at 180 and 246 ms were above chance but not significant after Bonferroni correction.)

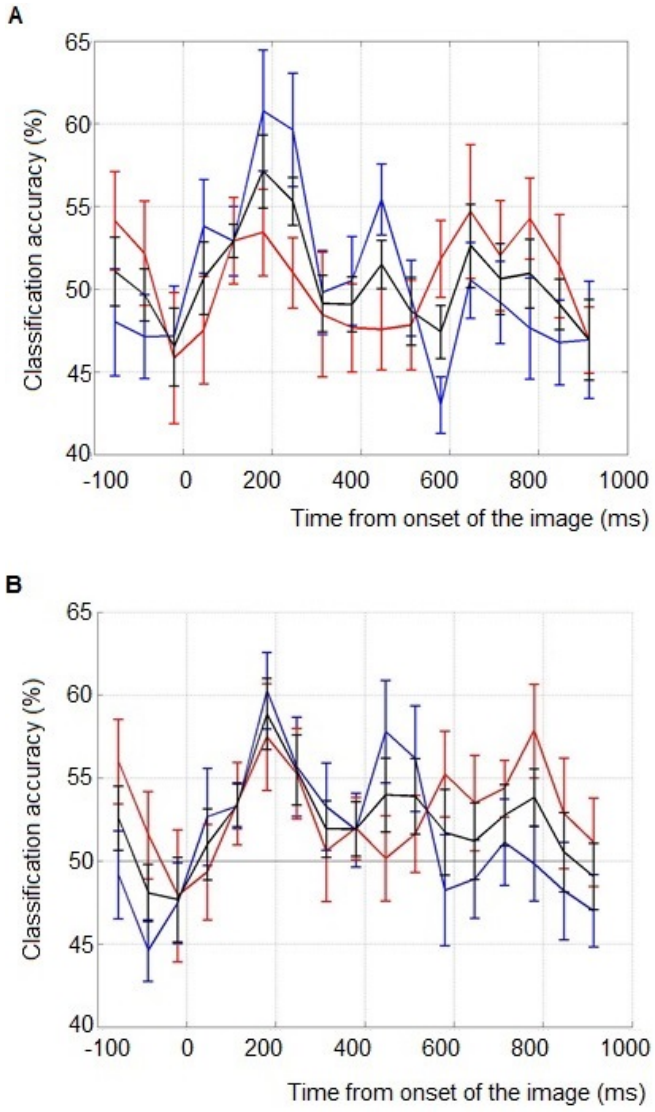

Figure 2.3: Feature reduction with univariate t-tests Average MLP classifier accuracy in classifying faces versus scenes, the error bars indicate standard error of the mean (A) Decoding using 12-90 Hz data showed significant classification at 180 and 246 ms (B) and decoding using 8-45 Hz was significant at 180 ms. The classification accuracy for faces is in blue, for scenes is in red and for the average classification is in black.

As an alternative feature reduction, I used thresholding. MLP Pattern classification performance using 12-90 Hz frequency range was above 50% at 113 ms (uncorrected $P = 0.011$), 180 ms (uncorrected $P = 0.006$) and 246 ms (uncorrected $P = 0.0034$). However, the classification was not significantly better than chance ($P = 0.003$; Figure 2.3 A). The classifications at 180 and 246 ms using thresholding feature reduction were not significantly better than using PCA (180ms : $P = 0.276$ and 246 ms: $P = 0.206$).

However, limiting the frequency range to 8 to 45 Hz improved the classification. Averaged accuracy of MLP classifier on this data set was significantly better than chance at 180 ms after the onset ($P < 0.001$). Thus, including the alpha frequency range improved classification performance (Figure 2.3 B; although two sample t-test did not show significantly better classification at 180 ms using 8 to 45 Hz than 12 to 90 Hz frequency range:$P = 0.452$).

2.4.3 Classification algorithms

The 8 to 45 Hz data was decoded with three classification algorithms. The SB classifiers showed above chance classification at 180 ms (uncorrected $P = 0.012$), 247 ms (uncorrected $P = 0.014$), and 780 ms (uncorrected $P = 0.023$). The SVM classifiers showed above chance classifications at 113 ms (uncorrected $P = 0.021$), 180 ms (uncorrected $P < 0.001$), and 247 ms (uncorrected $P = 0.028$). And the MLP classifiers showed above chance classifications at 180 ms (uncorrected $P < 0.001$), 247 ms (uncorrected $P = 0.030$) and 780 ms (uncorrected $P = 0.021$). Figure 2.4 shows the performances.

Overall, only SVM and MLP at 180 ms significantly performed better than chance (Bonferroni corrected threshold is $P < 0.003$). One-way ANOVA comparing the per-

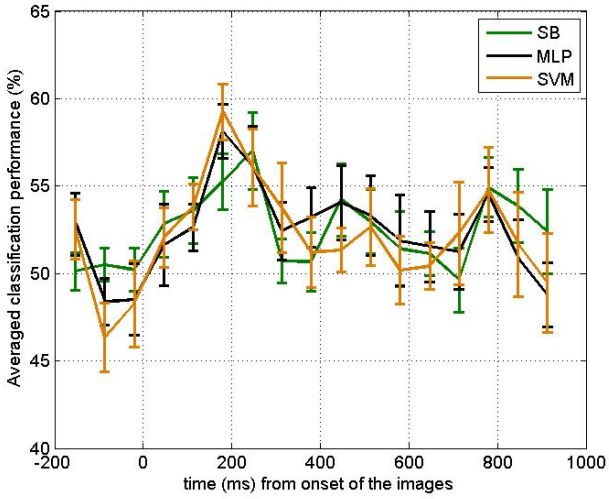

Figure 2.4: Cross-validation performance using SB, SVM, and MLP Average classifier accuracy, the error bars indicate standard error of the mean. The performance of SB is in green, SVM is in orange and MLP is in black. SVM and MLP at 180 ms can significantly classify better than chance (Bonferroni corrected threshold).

formance of classifier at 180 ms did not show a main effect of classification algorithm ($F(2, 27) = 1.53, P = 0.235$). However, the classification using MLP ($P = 0.047$) and SVM ($P = 0.024$) were significantly better than SB at 180ms; yet not significantly different from each other ($P = 0.34$).

2.5 Discussion

Through out this chapter I developed a pipeline to decode category-specific oscillatory representations at alpha/beta/gamma frequency range. Results suggested to z-normalize the data at each time point across the scalp rather than trial- or category-specific baseline correction, in order to avoid possible influence of baseline information at other time points or systematic bias toward categories. Also, results indicated that t-test feature reduction (without PCA feature reduction) leads to higher classification performance overall. That is about 5% more than using PCA. The advantage of t-test to PCA is possibly because the variation in representation of categories (scenes and faces) is too large in this data set.

In this chapter, I used time-frequency representation of data at 12-90 Hz and 8-45 Hz frequency range. I chose this frequency range to exclude well-known memory-related oscillatory activity in the theta frequency range and focus on the information representations in alpha/beta/gamma frequency (see also chapter 3). The results suggested that including lower frequencies improves the classification performance. In fact, new studies showed that including even lower frequency bands increases the classification performance yet further to about 90% (Carlson, Tovar, Alink, & Kriegeskorte, 2013; van de Nieuwenhuijzen et al., 2013).

In this chapter, I tested multiple pattern classification algorithms among which MLP and SVM (linear kernel) significantly classified categorical representations at about 180 ms from onset of the images. The classification performance was about 60% (SB: 58.18%, SVM: 59.23%, and MLP: 58.12%). Recent studies have improved the classification accuracy (at the same time window; about 180 ms) by including lower frequencies (Carlson et al., 2013). For example, Carlson et al. (2013) used PCA for feature reduction and LDA [2] classifier and used frequencies below 50 Hz to obtain about 88% accuracy. Other impressive results have used decoding in the source space. van de Nieuwenhuijzen et al. (2013) used an elastic net logistic regression algorithm and all the frequencies below 150 Hz to obtain up to 94% accuracy. In this chapter, however, the aim was to develop a pattern classifier to decode category-specific representation at alpha/beta/gamma frequency range only.

[2]Linear Discriminative Analysis is a liner model for classification. The linear discriminant function is obtained by a linear function of the input vector so that $y(x) = \sum_{m=1}^{M} w_m \phi_m(x)$ where $\phi_m(x)$ is the basis function and w is a weight vector. Therefore, the corresponding decision boundary is defined by the relation $y(x) = 0$, which is a $(M-1)$-dimensional hyperplane within the M-dimensional input space. From this boundary we can classify the new data points into the classes (Bishop, 2006c).

Chapter 3

Decoding replay of encoded information and timing of recollection

3.1 Precis

Long-term memories are linked to cortical representations of perceived events, but it is unclear which types of representations can later be recollected. Using Magnetoencephalography (MEG) based decoding I examined which brain activity patterns elicited during encoding are later replayed during recollection in the human brain. The results show that the recollection of images depicting faces and scenes is associated with a replay of neural representations that are formed at very early (180 ms) stages of encoding. This replay occurs quite rapidly, namely about 500 ms after the onset of a cue that prompts recollection and correlates with source memory accuracy. Thus, long-term memories are rapidly replayed during recollection and involve representations that were formed at very early stages of encoding. These findings indicate that very early representational information can be preserved in the memory

engram and can be faithfully and rapidly reinstated during recollection. These novel insights into the nature of the memory engram provide constraints for mechanistic models of long-term memory function [1].

3.2 Introduction

Recollection is associated with re-experiencing details of events, such as the scenery in which it took place or faces of individuals who were present (Tulving, 1985). There is now converging evidence that brain activity patterns that participated in representing aspects of these event characteristics during encoding can be later reinstated or *r*eplayed at retrieval; for a review see (Duzel et al., 2010). An intriguing puzzle in memory research is that cortical representations of event contents, such as faces, emerge very rapidly, within 200 ms (McCarthy, Puce, Belger, & Allison, 1999; Puce, Allison, & McCarthy, 1999; Fisch et al., 2009; Rossion & Caharel, 2011) whereas encoding processes in brain regions that are critical for recollection (i.e. the hippocampus and surrounding medial temporal areas) are initiated at 200 ms (Fell & Axmacher, 2011) and require several hundred milliseconds to unfold, as evidenced in invasive recordings of neural oscillations (Lega, Jacobs, & Kahana, 2012) and slow potentials (Fernndez et al., 1999; Axmacher, Lenz, Haupt, Elger, & Fell, 2010). These discrepancies in timing raise the question whether rapidly emerging cortical event representations formed at early stages of encoding are conserved in long-term memory and thus can be later replayed.

In order to capture the precise temporal evolution of neural representations during memory encoding and retrieval in the human brain, I used MEG-based multivariate

[1] This chapter derives in part from: "Replay of very early encoding representations during recollection" Jafarpour A, Fuentemilla L, Horner A, Penny W, Duzel E. (2014) *Journal of Neuroscience*; 34(1): 242-248

pattern classifiers (MVPCs) as outlined in chapter 2. It should be noted that methodological approaches based on univariate statistics are also suitable to detect neural reactivation as has recently been demonstrated with sophisticated experimental designs utilizing perceptual features of stimuli (such as flickering frequency) (Wimber, Maaß, Staudigl, Alan, & Hanslmayr, 2012).Wimber et al. (2012) have shown that the flickering frequency of stimuli presentation during encoding is detected at about 200 ms after onset of the memory cue. However, this rapid replay is considered to be unconscious and possibly reflecting implicit memory as subjects could not recollect the flickering information at retrieval.

For this study, healthy young adults were instructed to encode images of scenes and faces which were paired with words. Later, the words were used to probe image recollection. I trained MVPCs to decode oscillatory (8-45Hz) brain activity responses to images of faces and scenes during encoding when only the images were on the screen. MVPC analysis was performed every 66 ms, permitting us to capture the temporal evolution of neural representations. Then, I used classifiers which successfully classified the oscillatory activities into faces and scenes to detect the timing of replay of the same neural activity pattern at retrieval, when the word associated to the image was shown as a memory cue.

3.3 Methods and materials

3.3.1 Participants

Eleven right-handed healthy adults with normal or corrected vision participated in this experiment (6 female; mean 23 years old and SD of 2). All participants gave written informed consent to participate. The study was approved by the University

College London Research Ethics Committee for human-based research. All participants were financially compensated for their participation.

3.3.2 Stimuli and experimental design

The experiment contained six runs which consisted of two separate phases: the study (encoding) and the test (retrieval) phase. An arithmetic distraction task separated the two phases. In the study phase of each run, participants were required to memorize a set of 20 trial unique images associated with 20 trial unique words. All images were grey scaled and normalized to a mean grey value of 127 and SD of 75, of dimensions 300x300 pixels, and shown upon a grey background (grey value of 127) subtending approximately 6 degrees of horizontal and vertical visual angle. In each run, images were randomly selected from faces (5 female and 5 male) or scenes (5 indoor and 5 outdoor)(Figure 3.1A). And the paired words denoted either living (50%) or non-living (50%) objects with Kucera-Francis frequency (Kucera & Francis, 1967) of 20-24. Image-word associations were not semantically related and they were shown only once during encoding and were randomized across participants.

Participants were instructed to learn the association between the image and the word. For each association, scene or face images were presented for 2000 ms preceded and followed by a 1500 ms fixation period. Immediately thereafter, the same image appeared with the associated word in red for 3000 ms on top of it, which was followed by a living/non-living judgment about the word (responding with the index or middle finger of their right hand). After a random inter-trial interval of 1500, 2000 or 2500 ms the next image and image-word association was presented. An arithmetic task of 5 minutes separated the Study and Test phase to eliminate active rehearsal of the last image-words pairs studied in each run (Figure 3.1B).

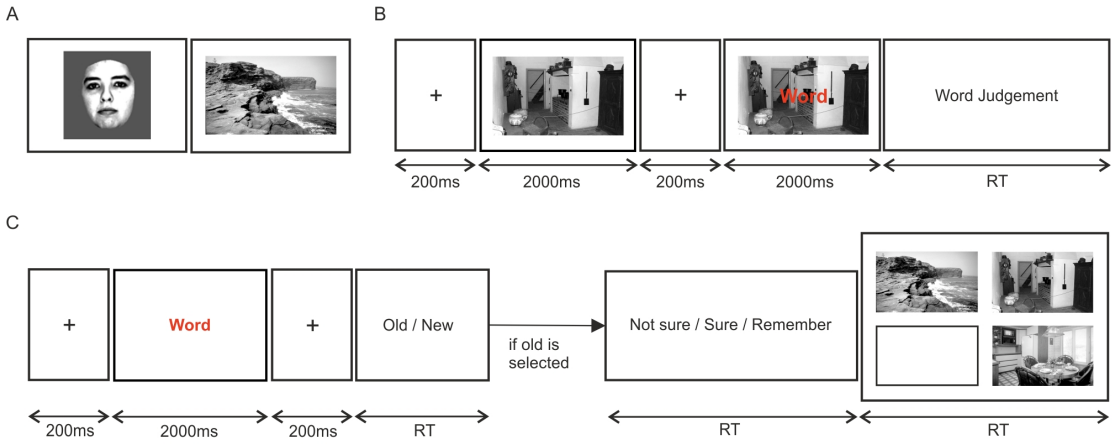

Figure 3.1: Schema of experimental paradigm (A) Samples of stimuli used in this experiment, a face and a scene. (B) Schema of experiment pipeline during encoding. The yellow frame in here demonstrates the epoch during encoding in which I trained and tested the classifier for decoding image category (faces or scenes) based on MEG oscillatory activity (8 to 45 Hz) at different time bins. (C) Schema of experimental pipeline during retrieval (when the word recognition old/new response is 'old'). The blue frame indicates the epoch in which I tested the replay of associated image after onset of the cue. The classifier which successfully categorized faces and scenes during encoding was then used to detect reactivation of categorical representations during encoding and during retrieval, when memory for the image was cued by the paired word.

In the test phase, a word (in red) was presented for 2000 ms. Afterwards, when an 'Old/New' question appeared on the screen, participants were required to judge whether the word was presented in the previous study phase (old) or was experimentally novel (new) with the right index and middle finger respectively. In each run, 20 Old and 20 new words were presented in a randomized order. Thereafter, a confidence judgment task (2000ms) followed. Here, new judgments were followed by 'Sure/ Not sure' and old judgments were followed by 'Remember/ Sure/ Not sure'. Participants were instructed to make confidence judgments following old judgments with respect to their ability to recollect the image associated to that word at encoding. They responded 'not sure' when they did not have any memory for the associated image, 'sure' when they believed that they could recognize what image was associated to the word, and 'remember' when they have the associated image vividly in mind.

Each trial ended with a source memory (image-selection) test during which three images and an empty square were presented in the four corners of the screen. The three images always included the face/scene originally paired with the word and two familiar images (i.e., presented in the Study phase with different words) from the same category as the paired-image. Participants were required to select, given 3000 ms time limit, which of these images was paired with the word or had the opportunity to select the empty square if they could not identify the match. A random inter-trial interval of 1500, 2000 or 2500 ms preceded the beginning of the next trial. After the test phase, participants had a short rest period before the next run (Figure 3.1C).

3.3.3 MEG preprocessing

Data were preprocessed using Matlab 2009 and SPM8 (`http://www.fil.ion.ucl.ac.uk/spm/software/`). The main noise (49 to 51 Hz) was filtered out of the data. Then MEG single-trial epochs of -1000 ms to 2500 ms relative to the onset of the images on the screen, when the images were shown for the first time, in the study phase and the same time window relative to the onset of the memory cue in the testing phase were extracted and baseline corrected (subtraction by the average amplitude of the epoch, offset correction). Next, for ERF analysis, the signal was low-passed filtered (45 Hz).

Then, for the TF analysis at the retrieval, the signal from -150 to 900 ms from onset of the cue was transferred to the time-frequency domain using 5 cycle Morlet wavelets. Wavelets from 3 to 45 Hz (including 8 to 45 Hz and 3 to 8 Hz, theta frequency) in step of 1 Hz were used for the transformation. Then the signal was rescaled to log values and baseline corrected using -150 to 0 ms ($\log power_{epoch} - \log power_{baseline}$). For the second level TF analysis was done in two ways. First, I analysed the average time $\times$ frequency over all channels using 3 to 45 Hz and -100 to 800 ms from onset of the memory cue. And next, I selected the time windows in which the replay was detected using pattern classifier, and used the average power in that time window (400 to 550 ms from cue onset) and its adjacent time windows (250 to 400 ms and 550 to 700 ms) for second level scalp $\times$ frequency analysis.

3.3.4 Data preparation

In another analysis, the signals from individual trials were again transferred to the time-frequency domain (TF) using 5 cycle Morlet wavelets. 38 wavelets were used for this transformation, from 8 to 45 Hz in steps of 1 Hz, and the power of the TF signal

was calculated. This frequency range covered a broad range of frequencies without compromising temporal resolution too much by including lower frequencies. Also, a model based on information theory suggests that power in alpha/beta frequency range reflects information coding in long-term memory (Hanslmayr, Staudigl, & Fellner, 2012), and 30-40 Hz frequency range is suggested to include the physiognomic information of faces category (Gao et al., 2013). The TF transformed data was then down-sampled to 300Hz and normalized by z-scoring the power value at each time, frequency and channel across trials.

For three subjects one run out of six experimental runs were discarded. For two of the subjects, there was a problem with triggers during the first run. And for another subject one run was discarded as she did not follow the experimental instructions during the first run. Thus, for three of the subjects five runs of six runs were analysed.

3.3.5 Pattern classifier analysis

A Support Vector Machine (SVM) with a linear Kernel (Vapnik, 2000b) was used to classify the preprocessed MEG signals of face versus scene samples. A pattern classifier was trained on MEG TF responses elicited when images of scenes and faces were shown at encoding (when the face/scene was first displayed on the screen, without the associated word). There were 60 samples of faces and scenes for seven of the analysed subjects and 50 samples of faces and scenes for three subjects. I used an equal number of samples from each category (random selection) for training and I tested an equal number of the remaining samples from each category (by random selection). The classification accuracy reported here is the performance of the classifier averaged over categories (faces and scenes), subjects and cross-validation folds (see below). The classifiers were trained separately for each participant and time bin. I

used 13 time bins, each of duration 66 ms, and centred at -19, 46, 113, 180, 246, 313, 380, 446, 513, 580, 646, 713, and 780 ms relative to stimulus onset. Each of the classifiers used spectral power in 274 MEG channels and at 21 time points within each time bin. For each time bin there were therefore 218652 possible features (274 channels x 21 time points x 38 frequencies).

For each of the 13 pattern classifiers (i.e., time bin) 10-fold cross-validation was adopted for validating the accuracy of the trained model. Accordingly, 10 classification iterations were run and 10% of samples from each category were left out at each iteration for testing the accuracy of the classifier. Prior to training, in each cross validation iteration, a feature-selection step was conducted by performing a univariate statistical analysis across the training set (excluding the validation set) on spectral power at each frequency, time point and channel that constituted the features for the classifier. The testing dataset was never included in the feature selection step. Those features which were found to be significantly different between categories by a two-tailed paired Student's t-test ($P < 0.05$), were selected. This data-led process served to reduce the dimension of the pattern classification problem by 95%; the selected features will be discussed below. In each cross-validation iteration the model was used to predict the category of the left-out trials (i.e. test trials). The classification performance was calculated as the average across the cross-validation iterations.

Classification performance at encoding was further investigated as follows. Firstly, I tested whether the classification accuracy during encoding was relying on the event-related field (ERF) component. For each subject I averaged the (low-pass filtered 45 Hz) signal for ERF in each category and subtracted the average category-specific ERF from the signal in each trial. The signal was then preprocessed as mentioned above (exactly the same as for the original signal) and cross-validation was again

the same process as for the original data. Secondly, the time bins in which pattern classifiers performed significantly above chance (in the main analysis) after multiple comparison correction were selected. Then, classifiers were trained on all trials from that encoding time bin. The trained classifiers were then used to classify all time bins during encoding. This analysis was performed to assess if the spatio-temporal frequency patterns that consistently contributed to classification at a specific time bin (e.g., 180 ms) were repeated at other time bins during encoding.

I next analysed the retrieval data in a similar fashion. I first selected time bins from encoding that showed significant classification performance in the initial cross-validation analysis. I trained classifiers for each time bin using all the trials for those encoding bins and tested on each time bin at retrieval, where memory for the images was cued with the associated word. Testing was performed at 13 separate time bins: -19, 46, 113, 180, 246, 313, 380, 446, 513, 580, 646, 713, and 780 ms from onset of the memory cue (the same time bins used in the encoding phase). The classification accuracy was calculated in relation to the category of the paired image (i.e., the image that the participant should have successfully retrieved). I studied retrieval in two steps. Firstly, I looked at replay in all the trials when the words were recognized correctly as *old* (recognition hits). In the second step, I analysed the trials in which the image associated with the word was selected correctly (source memory hits; recollection).

Between-subject (second-level) analysis of classification accuracy was implemented using SPM8 for MEG data. To test the accuracy of the classifiers against chance (i.e., 50%) I used a one sample t-test with a correction for multiple comparisons (family-wise error; FWE) using random field theory (RFT) implemented in SPM8 (Kilner, Kiebel, & Friston, 2005; Litvak et al., 2011). As a standard precedure in

neuroimaging, I made inferences using a cluster-level threshold. RFT procedure adjusts the p-value statistics that are function of number of time points (classification repetition in here). Such adjustment is similar to Bonferroni correction. However, Bonferroni correction is suitable for data sets which are independent at each repetition (or data point); whereas, time-frequency data is naturally not independent of adjacent time points and RFT is more suitable for multiple comparison correction (Kilner et al., 2005). To avoid numerical problems (e.g. infinite z-scores) in the input data for second level analysis in SPM8, I changed any 100% and 0% classification accuracies to 99.9% and 0.01% respectively (z-scores of which are 3 and -3).

Cluster-level family wise error corrected P value was used to examine the classification accuracy during encoding and retrieval of recognition hits trials. Based on decoding replay at retrieval, I studied the replay in selected time window at retrieval with source memory hits. Hence, in this case there were two time points of interest and I used the conservative Bonferroni corrected alpha level for t-test.

3.3.6 Time-Frequency analysis

For post-hoc classical univariate TF analysis at retrieval, 5 cycle Morlet wavelets from 3 to 45 Hz frequency range in steps of 1 Hz were used; similar to the preprocessing steps used for pattern classification. The power was transformed then to logarithmic scale and baseline-corrected by the average power in a -150 ms to 0 ms time window relative to onset of the word cue. For the second level analysis we used paired t-tests in SPM. SPM employs a family-wise error corrected statistical threshold (set at $P < 0.05$) for extracting the significance of statistical results (Litvak et al., 2011).

In the second level analysis, we assessed spectral power differences (in 3-45 Hz) at the time window during which MVPA indicated memory replay, 400-550 ms. The averaged power over 400-550 ms time window was calculated for each frequency and channel and then compared between hit and correct rejections (CR). A similar analysis was done for 250-400 ms and 550-700 ms time windows, which are the adjacent time windows to 400-550 ms and have the same time length. In the final step of the second level analysis, we compared the power (in the time windows of interest) between source hits and recognition misses using the same time windows.

3.4 Results

3.4.1 Behavioural data

Behaviourally, participants recognized the words at test equally well (corrected hit rate: M = 87.96% and SD = 5.05%) irrespective of the category of the paired image (hit rate for words associated with faces: M = 87.21% and SD = 6.85% and hit rate for words associated with scenes: M = 88.84% and SD = 5.72%, paired-sample t-test: $t(9) = -0.599, P = 0.563$). However, their source memory for scenes (hit rate: M = 80.11% and SD = 11.83%) was better than for faces (hit rate: M = 67.17% and SD = 16.82%; $t(9) = 2.91, P = 0.017$; Figure 3.2 A).

Repeated measures ANOVA, conducted on the source memory test as a function of source memory confidence and image category, revealed no main effect of image category ($F(1, 9) = 1.21, P = 0.276$); but there were significant effects of confidence level ($F(2, 18) = 31.46, P < 0.001$) and a confidence x image category effect ($F(2, 18) = 6.96, P = 0.002$). Post-hoc paired-sample t-tests indicated that subjects had more confidence (*r*emember) in selecting the correct scenes than correct faces

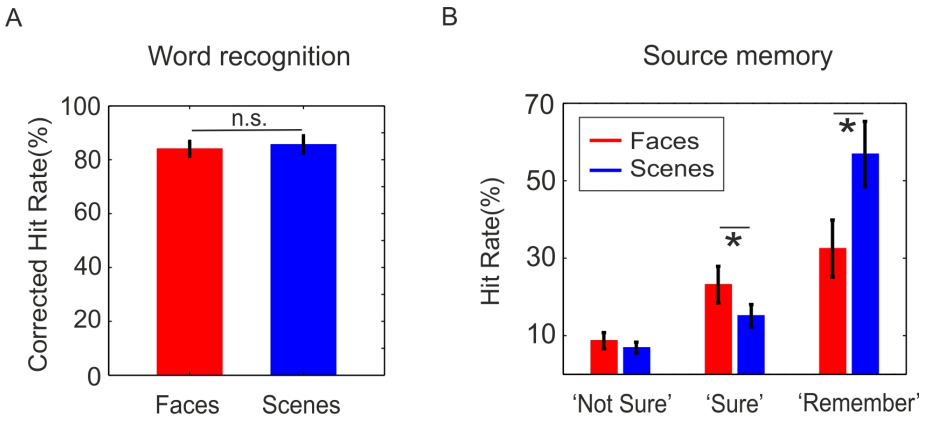

Figure 3.2: Behavioural results of the associative recognition experiment (A) Plot shows correct recognition performance (corrected hit rate) for words paired with scenes (in blue) and faces (in red). (B) Plot shows proportion of source memory hit rate for images of faces or scenes, recollected with low *n*ot sure, high *s*ure and very high *r*emember confidence. ($P < 0.05$, n.s. = not significant ($P > 0.05$), error bars represent SEM)

(for faces: M = 33.19%, SD = 22.76%, and for scenes: M = 57.40%, SD = 19.44%; $t(9) = 5.97, P < 0.001$) and they said *s*ure more frequently for correctly selecting faces (for faces: M = 25.85%, SD = 14.32%, and for scenes: M = 15.23%, SD = 9.29%; $t(9) = 4.77, P = 0.001$), but accuracy did not significantly differ in selecting the correct image when they were *n*ot sure (for faces: M = 8.12%, SD = 5.90%, and for scenes: M = 7.48%, SD = 5.00%; $t(9) = 0.48, P = 0.642$; Figure 3.2 B).

3.4.2 ERFs

The ERF analysis was done on two epochs, one during encoding and one at retrieval. During encoding, the signals elicited by onset of the images were averaged for each category, faces and scenes, for each subject. The group level analysis was performed then for studying any categorical-specific ERFs. The results revealed that there were significant differences (uncorrected $P < 0.001$) between ERFs of faces and scenes at right and left temporal channels Figure 3.3A at about 190 ms (M170; J. Liu et al., 2002; Gao et al., 2013). (ERF of the signal from the channels with most pronounced ERF differences during encoding are demonstrated in Figures 3.3C and 3.3E.)

At retrieval, the ERF signals elicited by the onset of the words were grouped by the associated image in order to study any categorical specific ERFs. The result indicated no significant difference between ERF of words associated with faces and those associated with scenes (uncorrected $P < 0.001$). (The distribution of ERF differences at the timing of replay (see below) is demonstrated in Figure 3.3B. And ERFs during retrieval from the channels, selected based on encoding ERFs, are demonstrated in Figure 3.3D and Figure 3.3F.)

3.4.3 Decoding

Cross-validation: decoding pattern of activity during encoding

MVPCs were used at different time bins during encoding to decode the emergence of category-specific neural activity elicited by the picture onset. At encoding, cross-validation analysis (Figure 3.5A, solid line) revealed significant above chance classification peaked at 180 ms after onset of the image (averaged classification accuracy = 59.20% at 180ms; peak-level $t(9) = 5.37$, cluster-level FWE-corrected $P = 0.001$, including 113, 180 (peak) and 246 ms).

According to cross-validation results, I trained the pattern classifier to classify the oscillatory pattern of activity at 180 ms using all the encoding trials. Hence, the feature reduction step was done across all trials to restrict the input of the classifier. I counted the percentage of selected features in different frequency bands and different channels. Figure 3.4 demonstrates the frequency-space distribution of features at 180 ms across different frequency bands, i.e. alpha (8-12 Hz), low Beta (13-19 Hz) and Beta/Gamma band (20-45 Hz), across subjects (see also chapter 5). Results indicated that equal number of features from each frequency bands was contributing. Furthermore, the distribution of features shows that the similar channels as those in the ERF analysis were included in the input features of the 180 ms classifier.

I next asked whether the significant classification at 180 ms was driven by the event-related M170 (J. Liu et al., 2002; Gao et al., 2013). Once we subtracted the mean category-specific ERFs from each individual trial, the subsequent classification analysis still revealed a significant above chance classification at 180 and 246 ms after image onset (average classification accuracy = 56.08% at 180 ms; peak-level $t(9) = 2.90$, cluster-level uncorrected $P = 0.009$, including 180 and 246 (peak) ms).This result suggests that the classification at 180 ms is not primarily driven

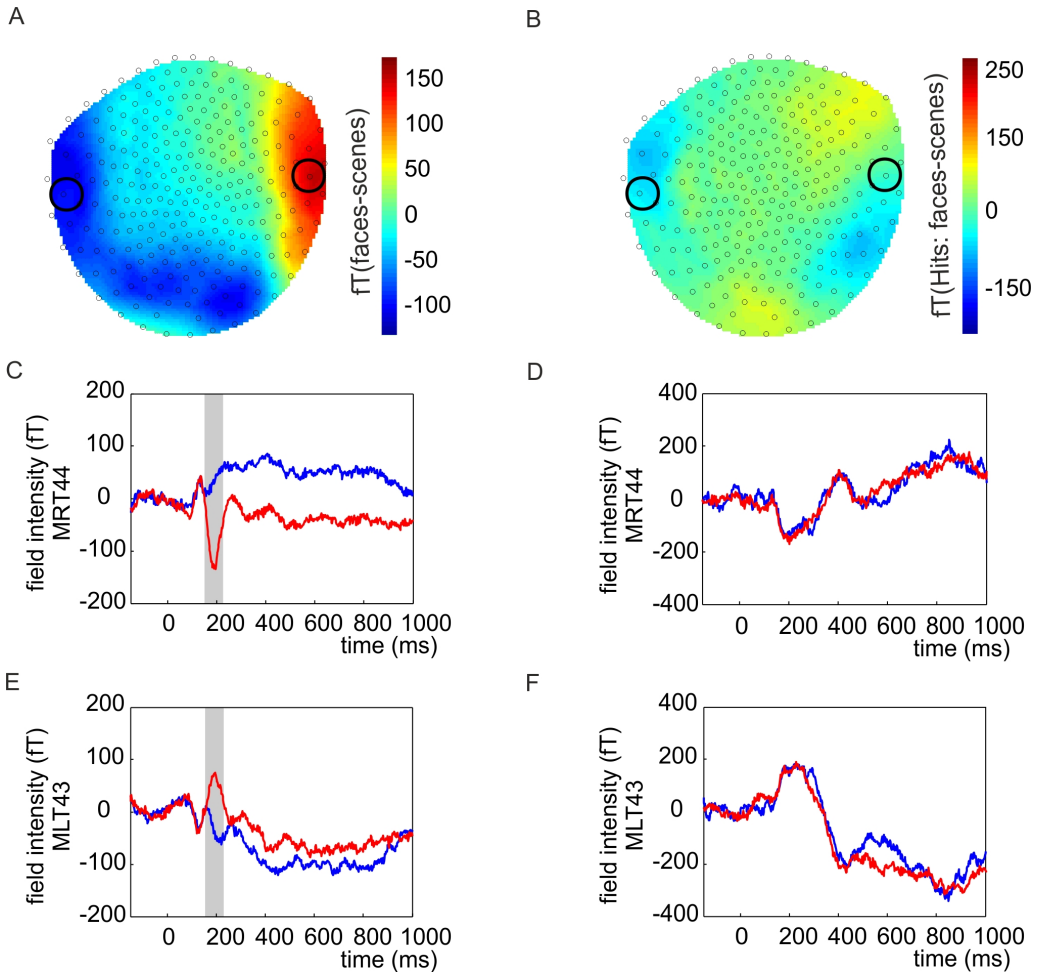

Figure 3.3: ERF responses at encoding (panels A, C, and E) **and at retrieval** (panels B, D, and F) (A) demonstrates the topographical distribution of difference signal intensity in ERF of faces and scenes at 185 ms. (C and E) show the ERF, M170, component at left and right inferior temporal channels (marked by dark circle in A) where the component is most pronounced. ERF for faces is in red and for scenes is in blue. The grey areas are where the difference between ERFs is significant ($P < 0.001$ uncorrected). (C) is the plot of ERF in right inferior temporal channel (MRT44) and (E) is in left inferior temporal channel (MLT43). (B) demonstrates the topographical distribution of difference signal intensity in ERF elicited by onset of the word associated with faces and scenes at 513 ms (when the replay is decoded by the classifier). Here we included the words which were correctly recognized as old (Hits). (D and F) show the event related responses to the word onset at memory test (faces in red and scenes in blue) in the same sensors as selected for encoding.

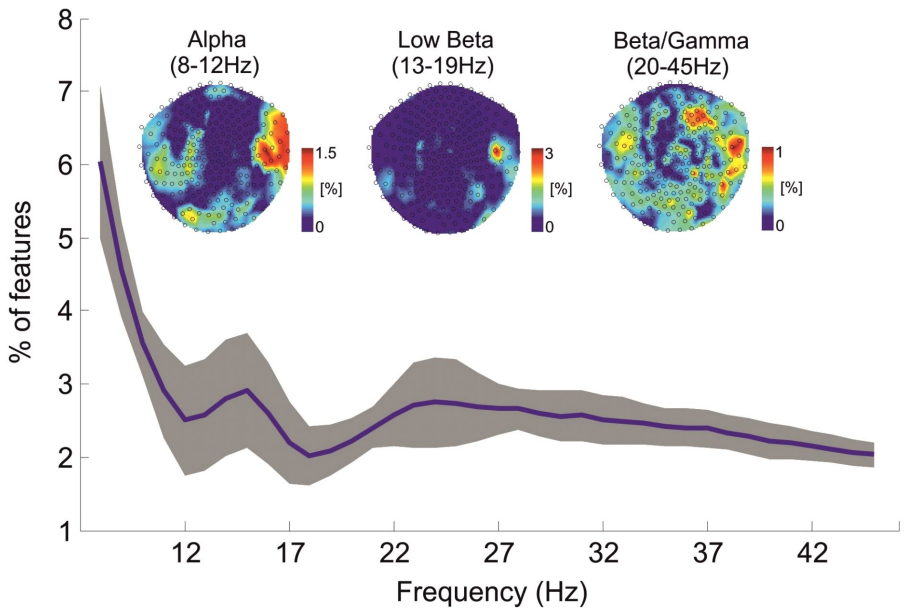

Figure 3.4: Distribution of the input features for 180 ms classifier The plot indicates the average percentage of features from each frequency across subjects. The grey area indicates the SEM of this distribution. Also each topographical plot indicates the percentage of features from the specific frequency band (Alpha (8-12 Hz) in the left, low Beta (13-19 Hz) in the middle and Beta/Gamma band (20-45 Hz) in the right plot). The color bar shows the percentage of features in each channel-frequency band.

by any category-specific ERF response (Figure 3.5A, dotted line). I note, however, that this classification analysis was marginally significant given my stringent FWE-corrected threshold, perhaps suggesting that the ERF component did contribute, albeit minimally, to classifier performance in my main encoding analysis.

Studying replay of the pattern of activity emerged at 180 ms during Encoding

I then tested whether the category-specific oscillatory patterns, which emerged at early time windows and peaked at 180 ms time bin, were replayed at any other time point within the first 800 ms of the encoding period. This was done by training the classifier on the oscillatory pattern at 180 ms time bin and testing during other encoding time points. As a result the 180 ms pattern was detected at the early time window during encoding (peak-level $t(9) = 11.74$, cluster-level FWE-corrected $P < 0.001$ including 46, 113 (peak), 246, 313 ms; Figure 3.5B). Therefore, correct classification rapidly dropped before and after the early time cluster. This suggests that face and scene-related neural representations at early time bins did not re-emerge at later time bins during the encoding period that I analysed. In sum, I found a category-specific oscillatory pattern at 180 ms that was not replayed at later time-points and was not primarily driven by a category-specific event-related response.

Decoding replay during retrieval and prediction of behavioural response

Next, I sought to investigate whether neural patterns identified at 180 ms during encoding were replayed during retrieval, when the memory was cued by the associated word. For this analysis, I used all the trials in which participants correctly

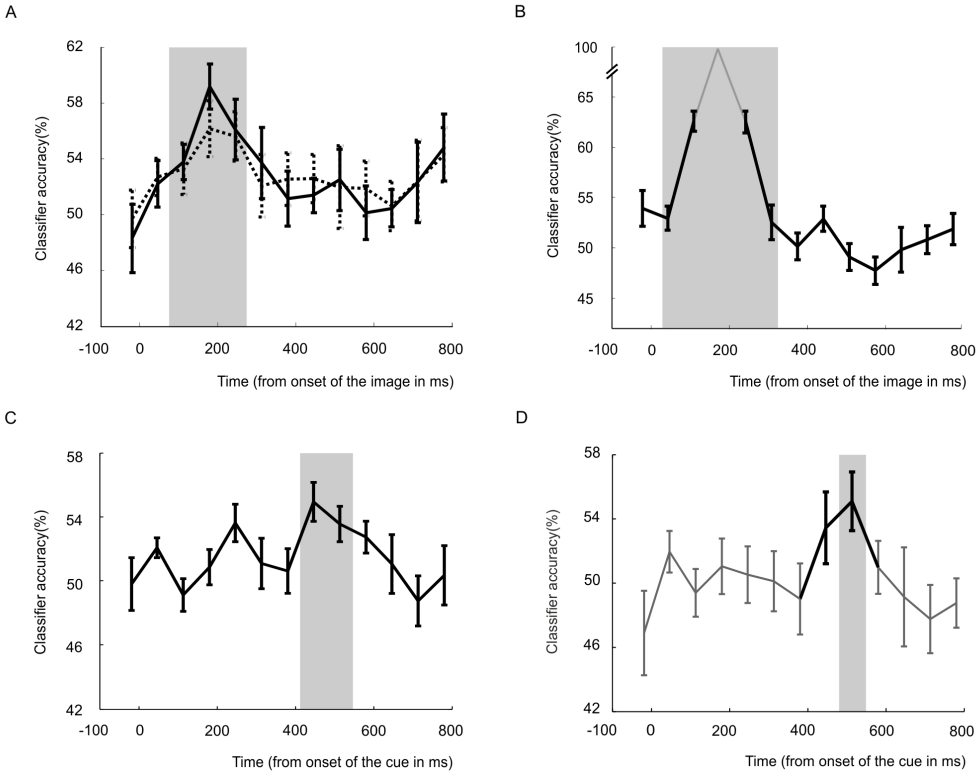

Figure 3.5: Category specific representations during encoding and retrieval. (A) The cross-validated accuracy of separate (for each time bin) pattern classifiers decoding faces and scenes during encoding (solid line). A rapidly (at 113 to 246 ms and peak at 180 ms) emergent pattern classifier decoded faces and scenes. Dotted lines show the cross-validated accuracy for decoding the signal from which the average category-specific ERF was subtracted (at 118 and 246 ms uncorrected $P = 0.009$ and cluster-level FWE-corrected $P = 0.061$). (B) The performance of 180 ms classifier in decoding other stages of encoding (significant only at the immediately adjacent time bins). 0 ms in (A) and (B) correspond to the onset of the images (face or scene) during encoding. (C) The same 180 ms classifier from the encoding period showed significant replay of associated image information at 446 to 513 ms from onset of the cue, *o*ld words, during correct word recognition. (D) In trials where the associated image is also correctly identified (recollected), replay is detected at 513 ms after onset of the cue. 0ms in (C) and (D) correspond to the word onset during retrieval. The time bins with significant classifications, multiple comparisons corrected $P < 0.05$, are highlighted in grey, see methods and results sections for details; error bars represent SEM. In (B) and (D) classification accuracy only from time points depicted with black lines were considered in second level analyses.

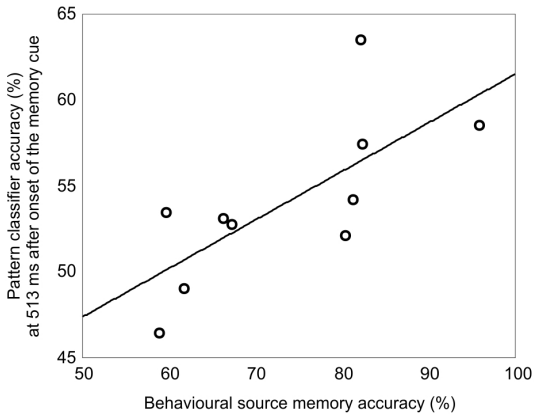

Figure 3.6: Classification accuracy correlates with source memory accuracy. Classification accuracy in decoding faces and scenes at 513 ms after onset of the memory cue (an old word) correlated positively with behavioural accuracy in source memory ($r = 0.73$ and $P = 0.017$). In this analysis, all trials were considered in which the word was correctly recognized as old (recognition hits). Each circle represents a participant.

recognized the word cue (i.e., Hits; averaged number of trials across subjects = 112 and SD = 9). The decoding revealed significant classification of oscillatory patterns elicited by the onset of words to the images associated with that word at 446 to 513 ms (peak-level $t(9) = 3.06$, cluster-level FWE-corrected $P = 0.022$ including 446 (peak) and 513 ms; Figure 3.5C). I therefore saw the same category-specific 180 ms oscillatory pattern seen at encoding during retrieval at about 450 ms after word onset.

Finally, I used these two time windows (446 and 513 ms) for hypothesis-driven testing of classification accuracy only in those trials in which subjects correctly selected the associated image. This was done in order to identify if the replay at 446 and/or 513 ms is associated with recollection (averaged number of trials across subjects = 66 and STD = 19). Congruent with this notion, I found significant classification at 513 ms (peak-level $t(9) = 2.64$, Bonferroni corrected $P = 0.026$ for testing two time windows; Figure 3.5D) for recollected trials. Furthermore, when all correct word recognition trials (irrespective to the source memory responses) were considered, the classification accuracy at 513 ms ($r = 0.73$ and $P = 0.017$; but not at 446 ms, $r = -0.07$ and $P = 0.833$) was predictive of the individual behavioural accuracy in source memory test (Figure 3.6). This relationship between replay and source memory performance strongly suggests a link between category-specific replay and the ability to recollect the contextual details of a previous event.

Category-specific replay

I further investigated the replay of each category separately. During encoding the cross-validation results were the same for faces and the scenes (Figure 3.7A); however, this was not the case at retrieval. The results indicated that the face information significantly replay at 446 ms (corrected $P = 0.026$; Figure 3.7C in red) and the scene information significantly replay at 513 ms (corrected $P = 0.018$; Figure 3.7C in blue).

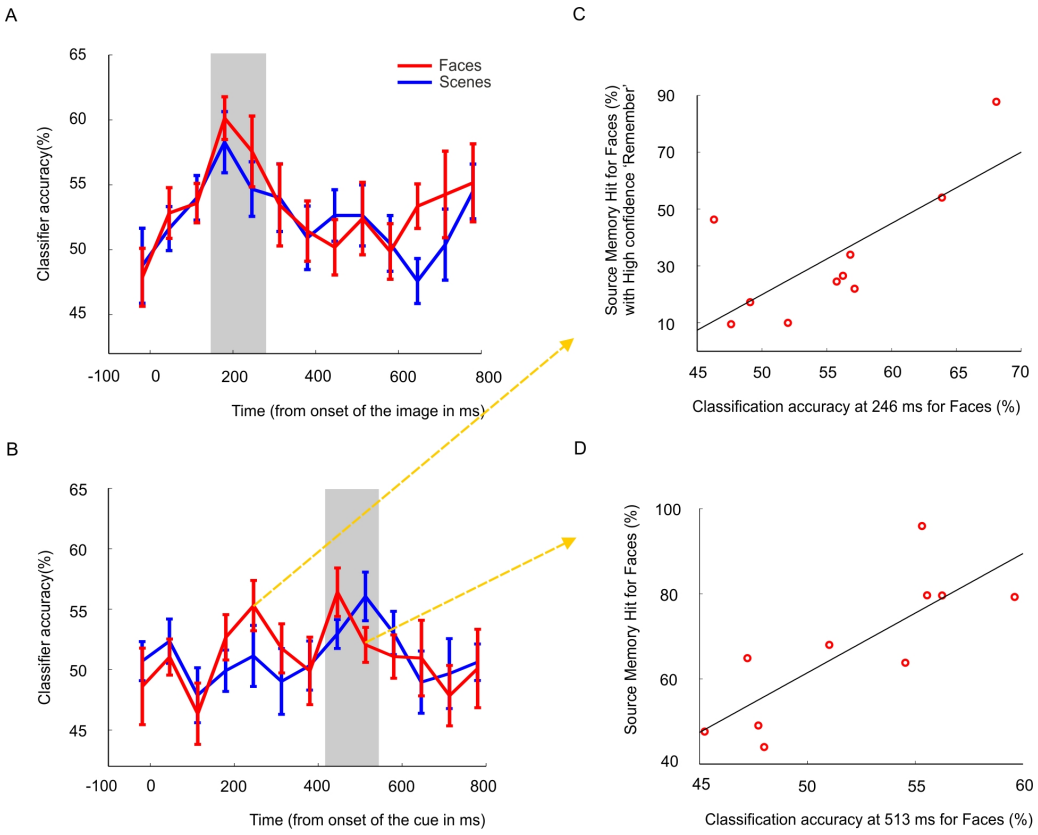

Figure 3.7: Face and scene specific representations during encoding and retrieval. (A) The cross-validated accuracy of separate (for each time bin) pattern classifiers decoding faces(in red) and scenes (in blue) during encoding. The classifier for faces was significant at 180 (peak) and 246 ms and for scenes was significant at 180 ms after onset of the image. 0 ms corresponds to the onset of the images (face or scene) during encoding. (B) The performance of 180 ms classifier from the encoding period showed significant replay of associated image information at 446 ms for faces and at 513 ms for scenes during correct word recognition. 0 ms in (B) corresponds to the word onset during retrieval. in (A) and (B) The time bins with significant classifications, multiple comparisons corrected $P < 0.05$, are highlighted in grey, see methods and results sections for details; error bars represent SEM. (C) Classification accuracy in decoding faces 246 ms after onset of the memory cue (an old word) correlated positively with behavioural confidence in source memory hits for faces ($r = 0.73$ and $P = 0.017$). (D) shows classification accuracy in decoding faces at 513 ms after onset of the memory cue positively correlated with behavioural source memory performance for faces ($r = 0.80$ and $P = 0.005$). In (C) and (D) each circle represents a participant.

Correlations between classification accuracy at retrieval (246, 446 and 513 ms) and behavioural responses (source memory performance, high confidence in source memory and category-bias) were tested. In order to correct for testing correlation between the behavioural measurements and multiple classifications, the Bonferroni corrected threshold of $P < 0.025$ was adopted. The results indicated that the replay of face information at 246 ms after the cue is predictive of subjects high confidence in selecting the correct associated face to the word ($r = 0.724$ and $P = 0.0177$; Figure 3.7C). Also the classification accuracy at 513 ms for faces (despite not being significant) is strongly predictive of subjects source memory for faces ($r = 0.804$ and $P = 0.005$; Figure 3.7D).

3.4.4 Time-frequency analysis

Group-level TF analysis revealed that at 400-550 ms (time window at which the MVPA indicated memory replay) there was a significant ($P < 0.05$ FWE corrected) theta power increase for hit trials compared to CRs that was maximal at left-temporal channels (Figure 3.8 A). A similar statistically significant ($P < 0.05$ FWE corrected) theta increase was also apparent in the adjacent time windows 250-400ms and 550-700ms. These results are congruent with previous studies contrasting recognition-hits and correct-rejections of word stimuli (Duzel et al., 2003, 2005). We also found that beta (23-25 Hz) power decreased for hit compared to CR trials over central and occipital channels at 400- 550 ms (Figure 3.8 B) and the following time window, 500-750 ms.

In a follow-up analysis we tested for power differences between source hits (correct picture selection) and recognition miss (misses) in the same three time windows. We

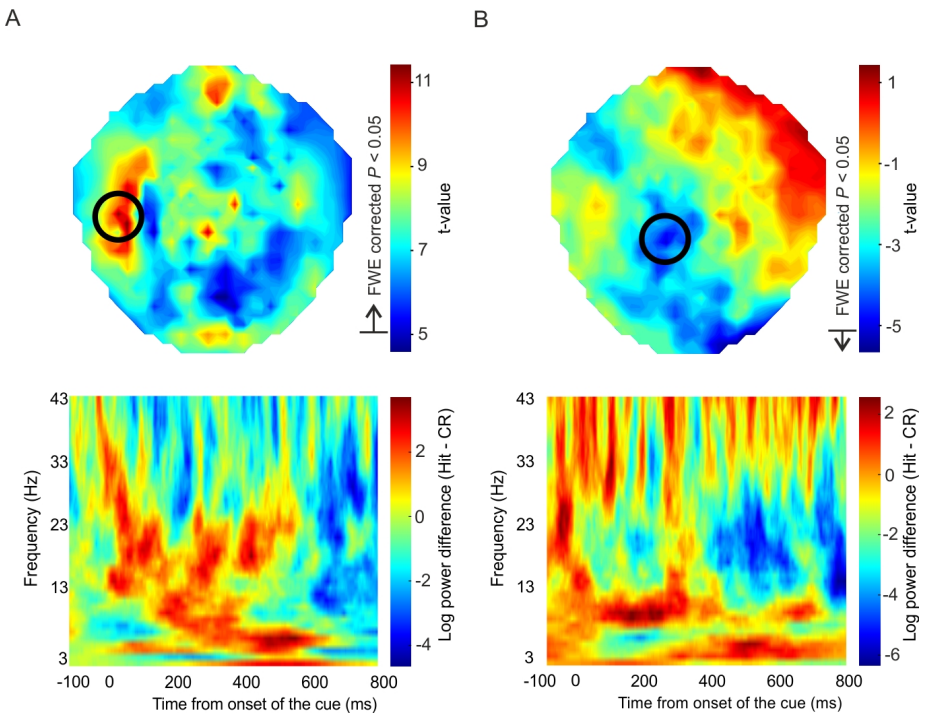

Figure 3.8: Power differences between Recognition Hits (Hits) and Correct Rejections (CRs) at 400-550 ms. (A) the top panel shows topographic distribution of t-value in Hits > CRs contrast at 3Hz, where the difference peaked at theta band. And the below panel shows the time-frequency representation of the log power differences between Hits and CRs (Hits - CRs), in a left temporal channel where the difference peaked (highlighted with the black circle). (B) top plot shows the t-value map for the difference at beta band, peaking at 23 Hz, And the below plot is the time-frequency representation of the log power differences at a central channel (marked with black circle) where the difference peaked. In (A) and (B) color-bars illustrate the range of t-value in the top plots and the arrows show the family-wise error (FWE) corrected t-value threshold (set at $P < 0.05$). The color-bar in other plots shows the range of log power difference between Hits and CRs.

found a significant (FWE-corrected $P < 0.05$) power decrease within the beta frequency range (13-25 Hz) at 400-550 ms and at 550-700 ms for source hits compared to recognition miss trials. This difference peaked over central channels similar to Osipova et al. (2006).

3.5 Discussion

My findings indicate that the category-specific neural representations of faces and scenes elicited selectively at the very early (before 200 ms after stimulus onset) stages of encoding are replayed during recollection. This replay of source information in this hippocampal dependent task (Horner et al., 2012) (Figure 3.1) occurs relatively rapidly, about 500 ms after the onset of the word cue and was predictive of behavioural accuracy in the source memory test (Figure 3.6). These results extend functional magnetic resonance imaging (fMRI) studies of single event memories that have shown that cortical activity patterns elicited during encoding reappear during subsequent memory retrieval (Kahn, Davachi, & Wagner, 2004; S. Polyn et al., 2005; A. Johnson, van der Meer, & Redish, 2007; B. A. Kuhl, Rissman, Chun, & Wagner, 2011; Ritchey, Wing, Labar, & Cabeza, 2012; Staresina, Henson, Kriegeskorte, & Alink, 2012). In these studies, the low temporal resolution of fMRI did not permit to determine whether the replayed patterns were established early or late during encoding and at which time bin(s) they were replayed during retrieval.

The time information obtained here addresses two major mechanistic possibilities regarding encoding. First, it shows that encoding is possible in the absence of prolonged maintenance of very early representations during encoding. The possibility that maintenance of information can aid encoding into long-term memory has been

recently suggested; for a review see (Hasselmo & Stern, 2006). While my results do not rule out such a possibility, they suggest that, if there is encoding-related maintenance, it does not involve replaying very early cortical representations (Figure 3.5B). Secondly, given the prolonged nature of encoding processes, the neural representations that are encoded and later replayed could be modified versions of the very early cortical representations. According to this possibility, memories are reconstructed during the later stages of encoding and consequently, the very early event representations cannot be reinstated during recollection. My results ruled out this possibility as they show that very early event representations can be reinstated during recollection.

In this study, I have investigated a special (albeit frequently studied) case of recollection in which memory content is composed of associations of single events (Figure 3.1). This may limit the generalizability of my findings to mechanisms underlying prolonged events such as those elicited during continuous spatial navigation (Hoffman & McNaughton, 2002; Fries, Fernndez, & Jensen, 2003) or movies (Gelbard-Sagiv, Mukamel, Harel, Malach, & Fried, 2008). Therefore, while I have positive evidence that very early representations can survive in long-term memory I cannot exclude the possibility that late representations are also replayed. Furthermore, while my results show replay associated with successful retrieval of scene/face associations (an index of recollection), I did not have a sufficient number of trials to assess whether replay is also absent in familiarity, i.e. when the items (words) were recognized but scene/face associations could not be recollected.

Two caveats should be considered with regards to the conclusions. Our MVPA is based on neural oscillations which are most likely to be largely cortical in origin. Our MVPA analyses are therefore likely detecting the reactivation of a cortical pattern rather than the retrieval mechanism that would necessarily precede (or trigger)

that reactivation (e.g., pattern completion in the hippocampus). As long as these retrieval mechanisms that trigger memory reactivation are inaccessible in relation to MVPA analyses, there remains uncertainty as to whether the reactivation that we observe is a direct consequence of retrieval processing (which we could also refer to as ecphory; Tulving et al., 1983) or results from additional post-retrieval processing and includes mental imagery. Furthermore, although (Horner et al., 2012) have shown a tight hippocampal dependence of this task, the link between replay and hippocampal function remains indirect as we cannot conclusively determine at which stage or time the hippocampus may have been involved.

The results suggest that pattern completion processes can lead to a reinstatement of the early neural representations of experienced events rather than changed or recoded versions of it. Thus, the memory engram (Dudai, 2012; X. Liu et al., 2012) must be sufficiently precise to enable the conservation of cortical event representations formed during very early stages of encoding. Encoding processes, in turn, despite their prolonged nature, appear capable of faithfully conserving initial representations of events without actively maintaining them in their early representation pattern.

Chapter 4

Decoding the content of visual working memory

4.1 Precis

In daily life, we constantly perceive sequences of events. A fundamental question is how the relative temporal order of the events is maintained in the human brain. To investigate this question we presented sequences of three visual items in a delay-match-to-sample task to 16 healthy young human adults and tested their memory for item-details and item-position in the sequence. A putative mechanism for maintaining temporal information of sequence relies on sequential replay of the information during the maintenance period. The goal of this study is to test this model. Thus far, I have analysed data from the encoding period and found a significant main effect of stimulus type on event-related fields (ERFs) at 100 to 200 ms after stimulus-onset and a main effect of temporal order of presentation at 400 to 500 ms after stimulus-onset. Furthermore, multivariate pattern classification (on 10 to 90 Hz frequency range) significantly decoded stimulus representations during encoding. As a future follow up study, the same classifiers will be used for decoding the content of working

memory during maintenance so as to test whether temporal order information is preserved during maintenance.

4.2 Introduction

The neural representation of an event is thought to include temporal information about when the event occurred. In fact a portion of neurons in CA1 of rodents codes for temporal information (time passed relative to the onset of the event), with these neurons known as "time cells" (Christopher, Lepage, Eden, & Eichenbaum, 2011). Specifically, time cell activity represents specific time bins during the delay between encoding and retrieval in working memory; the frequency of time cells activity is in theta range (Christopher et al., 2011).

It has been proposed that, in rodents, neural activity in the theta frequency range has a role in maintaining information about sequential order in working memory (Vertes, 2005). A physiological model of working memory maintenance suggests that there is repetitive replay of the information (represented by gamma oscillations) in theta oscillatory cycles(Lisman, 2010). For example, information about three stimuli, A B C, is maintained by replay of distributed and discrete neocortical representations of A, B and C, phase-locked to three consecutive phases of a theta cycle coordinated by the hippocampus (Sirota et al., 2008; Jensen & Colgin, 2007).

In line with this model, theta-locked reinstatement of gamma range representations of item-information has been shown to occur in monkeys (e.g. Siegel et al., 2009) and in humans during maintenance (e.g. Fuentemilla et al., 2010; Poch et al., 2011; Mizuhara & Yamaguchi, 2011). However, replay of neural patterns for the

maintenance of sequential information has not yet been characterized in humans. Specifically, the representation of sequences of stimuli and the impact of temporal order on the representation of stimuli are still unknowns.

To test this model, I used three stimuli from three district visual categories (faces, fruits and chairs) (Kriegeskorte et al., 2008) and I presented them in sequences of three items in random orders (first, second or third). MEG based pattern classifiers were used to decode the representation of stimuli in the 10 to 90 Hz frequency range. In a planned later step, the successful classifier will be used to decode the representations during maintenance.

4.3 Methods and materials

4.3.1 Participants

Sixteen right-handed healthy adults with normal or corrected vision participated in this experiment (8 female; on average 24 years old (SD=2)). All participants gave written informed consent and they were financially compensated for their participation. The study was approved by the University of London Research Ethics Committee for human-based research.

4.3.2 Stimuli and experimental design

A combination of a delay-match-to-sample and Stenberg memory tests was used. The experiment contained six runs. Each run consisted of a sequential presentation of three stimuli, a retention (delay) period and two probe tests. A run started with a fixation period for 4000 ms, then a presentation of a banana, a chair and a face in

a random sequential order. Each stimulus was presented for 500 ms, with a 500 ms inter-stimulus-interval (ISI). A 5500 ms retention period followed the presentation of the third item. Then a probe stimulus was presented to test for item memory (delay-match-to-sample), where subjects were required to select 'same' if the exact stimulus (type and perspective) was shown in the sequence and 'different' otherwise. Randomly, in half of the trials, the correct answer was 'same'. After the 'same or different' probe test, another stimulus was randomly selected from the sequence and presented to probe for temporal order memory (Stenberg test), where subjects selected 1, 2 or 3 according to the position of the stimulus in the sequence (Figure 4.1 A).

The stimuli were coloured images of a banana, a chair and a face photographed from three different perspectives - front-on, 60 degrees to the left, and 60 degrees to the right - shown upright on a white background, extending approximately 6 degrees of a horizontal and vertical visual angle. These three categories of visual stimuli were selected because they can be assumed to have relatively district spatial cortical representations (Kriegeskorte et al., 2008). This was done to improve decoding performance. Furthermore, multiple stimulus orientations were chosen to focus decoding on higher order and (possibly) semantic representations of stimuli rather than low-level perceptual responses.

In this experiment, unlike other MEG decoding studies (e.g. Newman & Norman, 2010; Fuentemilla et al., 2010; Jafarpour, Horner, Fuentemilla, Penny, & Duzel, 2013) which used trial-unique sets of stimuli from different categories, I used a single stimulus from each category and showed it from different perspectives (Figure 4.1 B). Selecting identical stimuli in this study was beneficial for reducing the categorical variance in representation and focusing on the representation of the temporal order

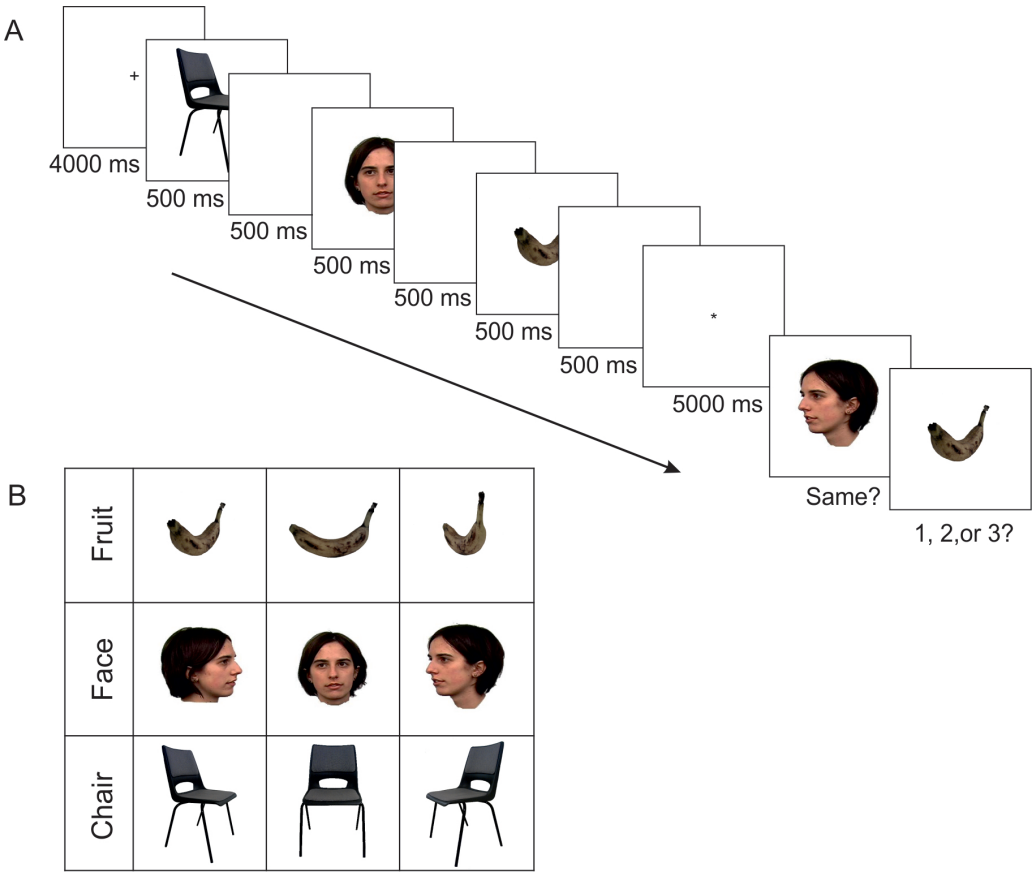

Figure 4.1: Schema of the working memory experimental paradigm (A) A trial started with a fixation point for 4000 ms; then three stimuli were presented sequentially, each for 500 ms and with 500 ms ISI. There was a 5500 ms retention period between the presentation and memory probe tests. The memory probe tests entailed a same or different judgment and a temporal order selection. (B) The stimuli used in this experiment: a fruit, a face and a chair from three points of view, 60° Left, 0° centre, 60° Right.

of stimuli.

Subjects were familiarized with the stimuli outside the MEG scanner. They also performed the experiment with feedback outside the scanner to ensure that they understood the experiment properly. There was no feedback given during the experiment inside the MEG scanner. I tested all the 162 ($3 \times 3 \times 3 \times 6$) possible sequential combinations of three stimuli.

4.3.3 MEG preprocessing and event-related responses

SPM12 was used for the analysis. The continuous MEG data were high-pass filtered at 0.2 Hz, low-pass filtered at 90 Hz and main noise (48 to 52 Hz) was filtered out using a fifth-order Butterworth filter. The MEG data obtained during encoding were epoched from -100 to 500 ms relative to the onset of the stimuli. Any epoch with field magnitudes greater than 1.5e-11 tesla in any channel was discarded as containing artefacts. Any such bad epoch was excluded from further analysis. The data were down-sampled from 600 Hz to 300 Hz.

ERFs elicited during stimulus perception were studied relative to the category of the stimulus (face, banana or chair) and the position of the stimulus in the sequence (first, second or third). A one-way repeated-measures ANOVA was used to test for the main effects of position and stimulus category, and t-tests were used for post-hoc comparisons. The statistical threshold was set at the cluster level family-wise-error corrected $P < 0.05$ and FWE-corrected $P < 0.05$.

In the next step, I identified the frequency range of activity which showed a main effect of category or position in the ERFs. Thus, I tested the TF representation of

activity within the time windows of the ERFs in 2 to 90 Hz (2 to 45 Hz in 1 Hz steps and 45 to 90 Hz in 2 Hz steps). The statistical threshold was set at the cluster level family-wise-error corrected $P < 0.05$ and uncorrected $P < 0.001$.

4.3.4 Eye-tracking

An Eyelink 2000 (SR reseach) high-resolution eye tracker was used and synchronized with the experimental protocol and MEG recordings. For technical reasons, the eye movement data were not acquired for one of the subjects. I recorded eye movements because the frequency of saccadic eye movements occurs in theta frequency (Carl, Açk, Konig, Engel, & Hipp, 2012) and theta frequency is also associated with various memory processes relevant to my research question. The frequency range of interest was 2 to 10 Hz (theta and lower alpha range) and the time window of interest was 400 to 500 ms after the stimuli onset.

In order to control for the potential effects of eye movements on the MEG data, I first tested for the main effect of experimental condition on the overall number of saccades in the time window of interest. To do so, the relative eye movement disposition, z, was calculated based on the horizontal and vertical movements: $z = \sqrt{(x^2 + y^2)}$, where x is the horizontal and y is the vertical position of the eye. It was then possible to identify the number of saccades having occured by counting the number of times that z was larger than 5 mm.

Next, the effect of experimental conditions on the frequency of eye movements was tested. In this regard, the horizontal and vertical tracks of eye position were transformed to the time-frequency domain, and an ANOVA was used to examine the main effects of the experimental conditions. In the follow up, to ensure that the

MEG event-related responses were not driven by eye movements, I first measured the variance of eye displacements in a time window of interest. Then I regressed out the power in the theta frequency which was explained by the eye movements (in the channel with the maximal event-related response). Then I tested for the main effects of interest on the residuals.

4.3.5 Decoding

For decoding, I selected the epoch when the first stimulus in a sequence was presented. There were two reasons for selecting this epoch. Firstly, from the participant's perspective, the content of this epoch was the least predictable, as the order of categories was random. Secondly, the working memory load in that epoch was one item, and therefore lower than for subsequent stimuli. For decoding, an equal number of samples from each stimulus was selected (about 54 samples per stimulus). Decoding was conducted twice; first using 10 to 45 Hz (in 1 Hz steps) and secondly using 45 to 90 Hz in 2 Hz steps. The TF data was not normalized. The classification was done at 30 time windows, each with 20 ms width, spanning -10 to 500 ms relative to stimulus onset. The power in each time window was averaged across the 20 ms time window (6 time points).

A SVM pattern classification algorithm was used and a 10 fold cross-validation approach was adopted (as explained in chapter 2). I used t-tests for feature reduction within each cross-validation iteration. The decoding was done in two ways. First, I trained classifiers to decode the representation of each stimulus versus a random selection of an equal number of stimuli from the other two types. Thus, if n is the number of banana images, the 'others set consisted of $n/2$ randomly selected samples from faces and $n/2$ randomly selected samples from chairs. Second, I trained

classifiers for pairwise decoding (face vs banana, face vs chair, and banana vs chair). Again, an equal number of samples from each stimulus was used. The decoding details were the same as in chapter 2 and chapter 3.

The second level analysis was the same as chapter 3. To test the accuracy of the classifiers against chance (i.e., 50%) I used a one sample t-test with a correction for multiple comparisons (FWE) using random field theory implemented in SPM (Kilner et al., 2005; Litvak et al., 2011). Cluster-level family wise error corrected P value was used to examine the classification accuracy.

4.4 Results

4.4.1 Behavioural data

On average, participants selected the correct position of the stimuli in the sequence (order test) in 93.06% of the trials (SD = 8.31%). There was no significant main effect of stimulus type in the test of order memory (one-way ANOVA: $F(2, 47) = 0.193$, $P = 0.825$) performance being similar for face (M = 93.47% , SD = 6.96%), banana (M = 93.68% , SD = 4.38%) and chair (M = 91.87%, SD = 9.91%). And the paired-samples t-test confirmed that participants remembered the position of stimuli irrespective of stimulus type (face versus chair: $t(15) = 1.044$, $P = 0.313$; banana versus chair: $t(15) = 1.701$, $P = 0.109$; face versus banana: $t(15) = -0.150$, $P = 0.883$; order test Figure 4.2 A).

The participants were also accurate at recognizing the correct orientation of stimuli (M = 94.25%, SD = 4.67%). However, paired-samples t-tests showed that they

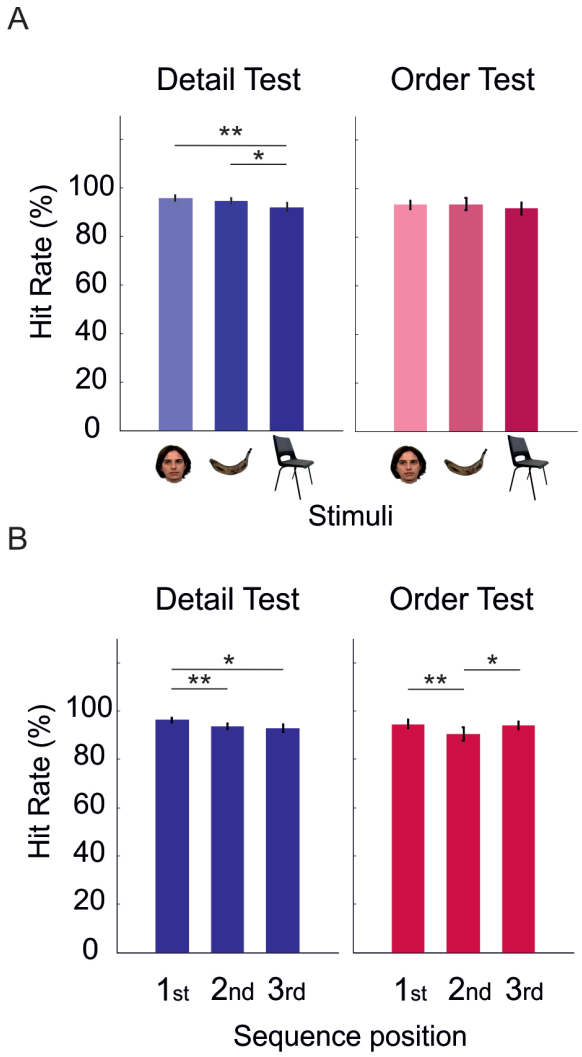

Figure 4.2: Behavioural results of the working memory experiment Behavioural accuracy in (Hit Rate) for detail test (same/different) and order test (1st/2nd/3rd) (A) relative to the stimulus category (face, banana or chair) and (B) relative to the stimulus position in the sequence (1st, 2nd or 3rd). ** denotes significant difference in performance ($P < 0.05$) and * denotes marginally significant difference ($P < 0.08$), error-bars depict SEM.

performed better for the face (M = 95.71% , SD = 4.69%) and the banana (M = 94.79% , SD = 4.38%) than the chair (M = 92.14% , SD = 7.06% ; face versus chair: $t(15) = 2.979$, $P = 0.009$; banana versus chair: $t(15) = 1.993$, $P = 0.065$; face versus banana: $t(15) = 0.763$, $P = 0.457$), although there was no significant main effect of stimuli type on memory performance (one-way ANOVA: $F(2, 47) = 0.193$, $P = 0.825$; Figure 4.2A).

I also tested the effect of sequence position on memory performance. Results showed no main effect of position on the detail probe test (same/different judgment; one-way ANOVA: $F(2, 47) = 0.969$, $P = 0.387$). Yet, paired-samples t-test showed a primacy effect for same/different test. In other words, participants were significantly better in same/different test, when the sample item was presented first in the sequence compared to second ($t(15) = 3.337$, $P = 0.004$) and marginally significantly better for the first than the third ($t(15) = 2.086$, $P = 0.054$), but there was no significant difference between second and third item in the sequence ($t(15) = 0.635$, $P = 0.535$; Figure 4.2 B).

The position of the tested stimuli in the sequence also did not have a significant effect on subjects accuracy in selecting the correct order (one-way ANOVA: $F(2, 47) = 1.787$, $P = 0.179$). Nevertheless, t-tests showed that participants memory was better in the order test for the first stimulus compared to the second ($t(15) = 2.382$, $P = 0.031$), and memory for second stimulus was marginally significantly worse than the third one ($t(15) = -1.887$, $P = 0.078$), but there was no significant difference in memory of the position of the first and last stimulus ($t(15) = 0.888$, $P = 0.388$; order test Figure 4.2A).

4.4.2 ERFs

Two-way factorial design tests were conducted. The factors were the stimulus category (face, chair, banana) and their order in the sequence (first, second, last). No significant interaction between the category and the position of stimuli in the sequence was found (using two-way ANOVA). However, significant main effects of the category and position were detected, as described below.

Stimulus category

The main effect of stimulus category was detected from 95 to 110 ms from onset of the stimuli at the occipital channels extending to left temporal channels (FWE-corrected $P < 0.001$; threshold was set at $F(2, 135) > 14.13$). The effect at 100 to 110 ms was also detected at right occipital channels (FWE-corrected $P = 0.004$). Also a left temporal cluster showed the effect at 170 to 180 ms (FWE-corrected $P = 0.018$). In addition, there was a main effect of stimulus-type at right occipital channels (where the effect was found at 100 to 110 ms) at 487 ms (FWE-corrected $P = 0.003$). Figure 4.3 shows the ERF components.

Post-hoc analyses using t-tests showed significant differences between ERFs for the face in comparison to the chair and banana. This difference entailed a positivity over left parietal channels at 180 ms (M170; FWE-corrected $P = 0.001$), a negativity at 120 ms over left temporal channels (FWE-corrected $P < 0.001$), at 170 ms (N170) over right temporal channels (FWE-corrected $P = 0.008$) and at 470 ms again over right temporal channels (FWE-corrected $P < 0.001$).

The difference between banana and other stimuli emerged as more positive waveforms at 200 ms over left temporal-posterior channels (FWE-corrected $P < 0.05$) and at 170 ms over parietal channels (FWE-corrected $P = 0.037$). Also ERFs for

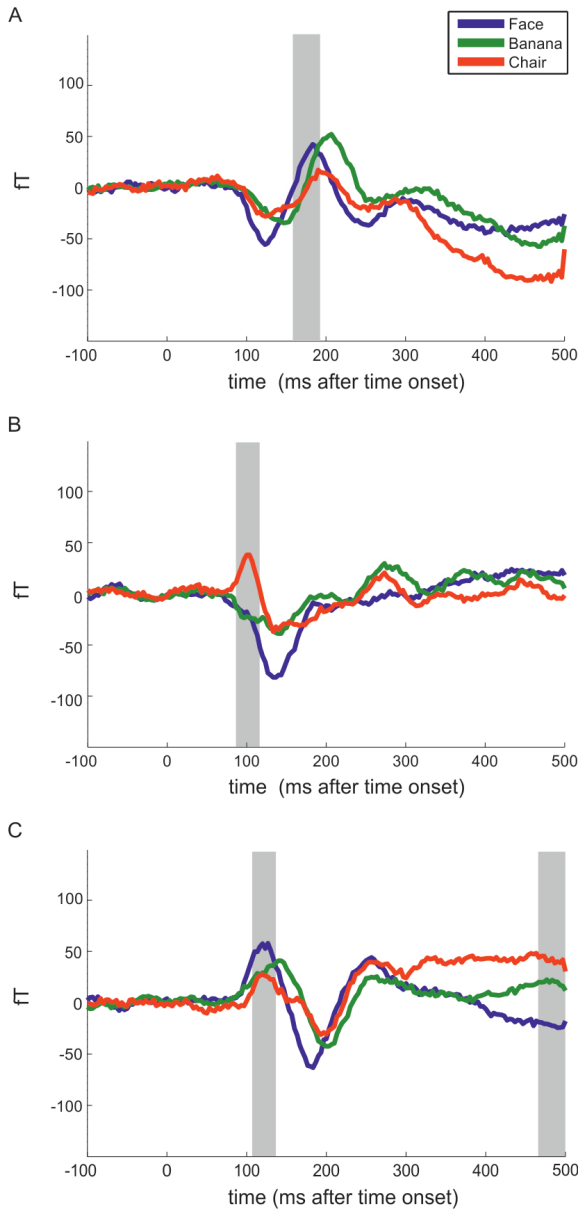

Figure 4.3: ERF of stimulus category during encoding (A) a left temporal channel (MLT53), (B) a left occipital channel (MLO33), and (C) a right temporal channel (MRT35). (A,B,C) The ERF for face is in blue, banana in green and chair in red. The grey areas show when the main effect of stimulus category was detected (FWE-corrected $P < 0.05$).

the chair differed significantly from the face and banana at 100 ms (positivity) over left occipital channels (FWE-corrected $P < 0.001$), (negatively) over right occipital channels (FWE-corrected $P < 0.001$) and (negatively) at 170 ms over left temporal channels (FWE-corrected $P = 0.023$).

Stimulus position in the sequence

The main effect of stimulus position was detected over the left central channels at 420 to 500 ms (FWE-corrected $P < 0.001$). The effect was also found over right frontal channels at 300 to 350 (FWE-corrected $P = 0.002$) and then at 400 to 500 ms (FWE-corrected $P < 0.001$). In addition, the main effect of stimulus position was found over left frontal channels at 460 to 500 ms (FWE-corrected $P < 0.001$). These results are similar to the P300 ERP component which is detected by EEG in similar timings and channels (Morgan, Klein, Boehm, Shapiro, & Linden, 2008) and which reflects working memory load (Figure 4.4).

The post-hoc t-tests indicated significant differences between the first stimulus presented in each sequence in comparison to the second and the last in the sequence: this difference was evident as a positivity over central channels at 450 ms (FWE-corrected $P < 0.001$), a negativity over right frontal channels at the end of the epoch (480 ms; FWE-corrected $P < 0.001$) as well as at 330 ms (FWE-corrected $P = 0.001$). No significant difference was found between ERFs of the second item and other items. And the ERFs of stimuli in the the third sequential position were different from others at 100 ms (positivity) over occipital channels (FWE-corrected $P < 0.001$) and (negativity) over left frontal central channels at the end of the epoch (470 ms, FWE-corrected $P < 0.001$; Figure 4.4).

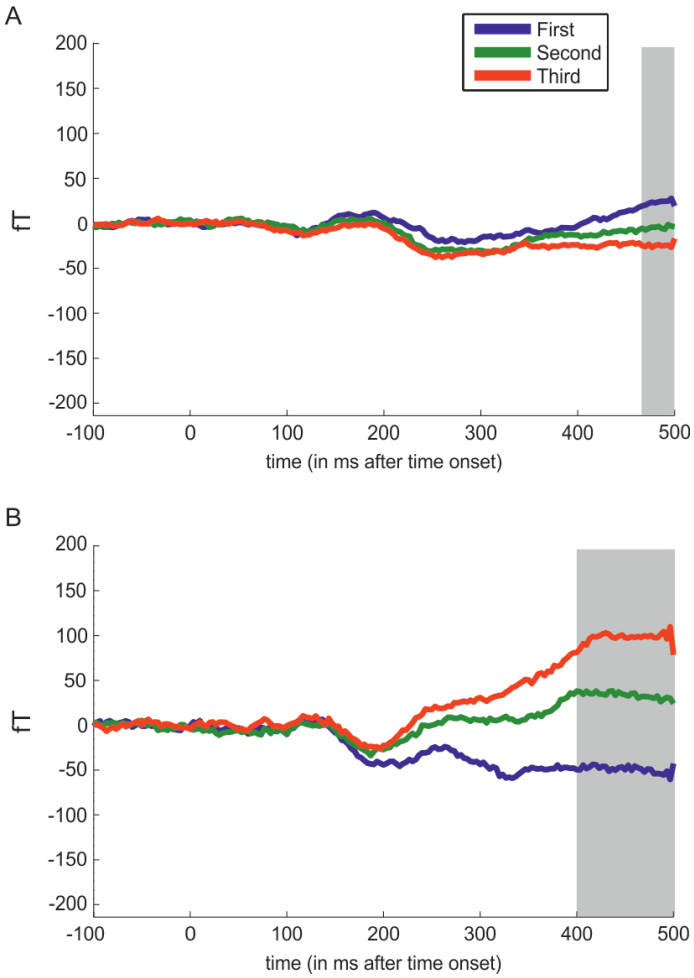

Figure 4.4: ERF of stimulus position in the sequence during encoding in (A) a left frontal channel (MLF55) and (B) a right frontal channel (MRF14). (A,B) The ERF for the first item in any sequence is depicted in blue, the second in green and the third in red. The grey areas show when the main effect of stimulus position was significant (FWE-corrected $P < 0.05$).

4.4.3 Time-frequency analysis

100 to 200 ms from onset of the stimuli

The ERF revealed the main effect of the category of stimuli at 100 to 200 ms from onset of the stimuli. The TF results showed an average effect of stimulus onset at 2 to 30 Hz (FWE-corrected $P < 0.001$) and 51 to 89 Hz (FWE-corrected $P < 0.001$) when considering all channels. The main effect of stimulus type, however, was detected only at 9 to 10 Hz over right temporal channels (FWE-corrected $P < 0.001$) and 22 to 30 Hz (peak at 27 Hz) over occipital channels (FWE-corrected $P < 0.001$; Figure 4.5 A). There was also a main effect of stimulus position in the sequence at 9 Hz over frontal channels (FWE-corrected $P = 0.013$). And no significant interaction was detected.

400 to 500 ms from onset of the stimuli

The ERFs showed a main effect of stimulus position at 400 to 500 ms from stimulus onset. The TF results showed an average effect of stimulus onset at 2 to 45 Hz (FWE-corrected $P < 0.001$) when considering all channels and 51 to 89 Hz over occipital and temporal channels (FWE-corrected $P < 0.001$). No significant main effect of stimulus type was detected. However, there were significant main effects of stimulus position at 3 to 8 Hz (peek at 5 Hz) over right temporal (FWE-corrected $P < 0.003$; Figure 4.5 B) and left temporal channels (FWE-corrected $P < 0.004$; Figure 4.5 B). No significant interaction was detected.

A post-hoc t-test showed that there was significantly more activity over frontal channels in the beta range (10 to 20 Hz, peaks at 13 Hz; FWE-corrected $P = 0.001$).

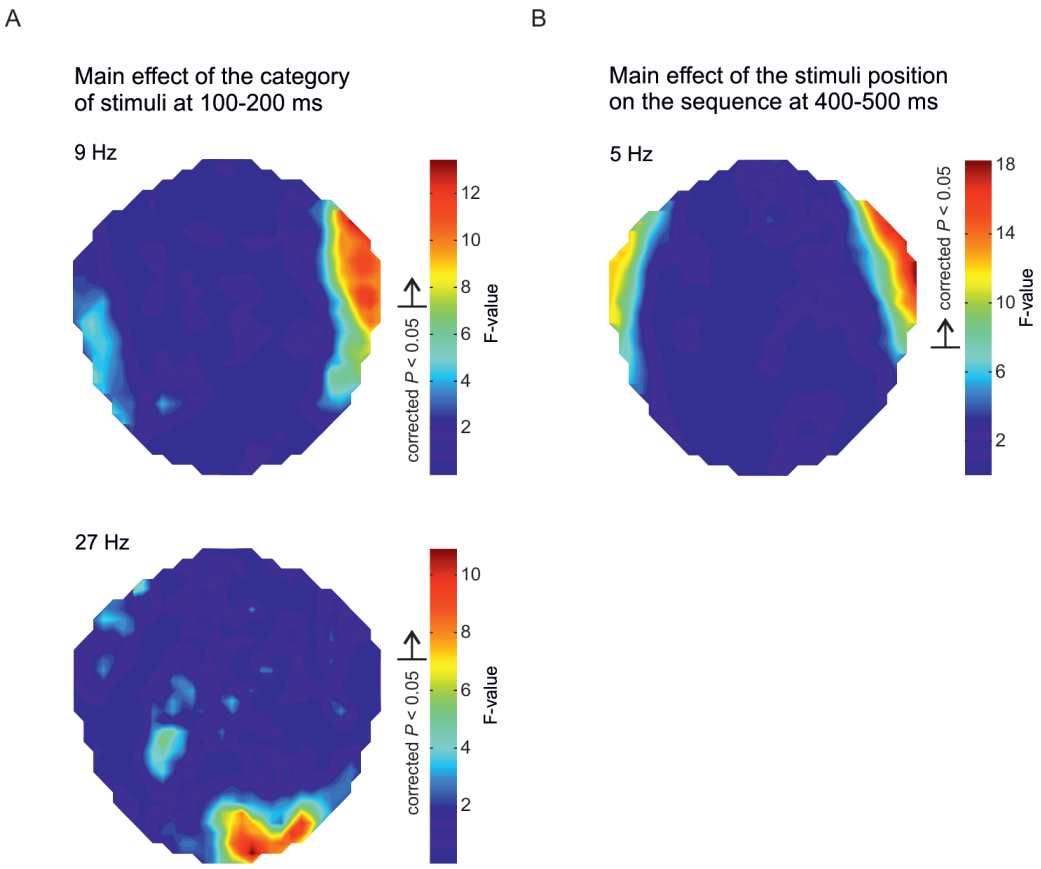

Figure 4.5: Time-frequency distribution of the main effects of stimulus category and their order of presentation at 100 to 200 ms and at 400 to 500 ms (A) The main effect of stimulus category at 100 to 200 ms from stimulus onset: The top plot shows the distribution of the main effect in the peak of the effect, 9 Hz in the alpha range (9 to 10 Hz). The plot below shows the peak of the main effect of stimulus type at 27 Hz - in beta range (22-30 Hz). (B) the main effect of stimulus position in the sequence at 400 to 500 ms: the peak effect at 5 Hz in the theta frequency range (3 to 8 Hz). In (A) and (B) colour-bars show the range of F-stats in each of the plots and the black like indicates the significance threshold ($F(2, 135) = 7.27$) which is FWE-corrected $P < 0.05$ at the cluster level.

The second stimulus in the sequence did not show any significant difference in power over 2 to 90 Hz in comparison to the first and last items. And the last stimulus in the sequence showed significantly stronger theta activity (3 to 9 Hz, peak at 5 Hz) over right temporal (FWE-corrected $P = 0.016$) and left temporal (FWE-corrected $P = 0.002$) channels.

4.4.4 Saccadic eye movements

I studied eye movements during encoding and compared them to the distribution of the effect of stimulus-position in theta frequency range. In the later time window (400 to 500 ms) theta effects looked like they could emerge from eye movement and in fact saccadic eye movements are usually in theta frequency. The percentage of trials with saccades at 400 to 500 ms of each presentation order (first, second, or third) was calculated. A significant main effect of stimulus position in the sequence and the number of saccades at 400 to 500 ms was detected ($F(2, 44) = 3.21$ and $P = 0.050$).

Post-hoc tests showed significantly more trials with saccades at the end of presentation of the last item ($M = 9.26\%$, $SD = 10.22\%$) compared to the first ($M = 3.33\%$, $SD = 4.21\%$, $t(14) = 2.70$ and $P = 0.017$) and the second ($M = 5.10\%$, $SD = 7.37\%$, $t(14) = 3.07$ and $P = 0.008$) items (Figure 4.6). (No significant effect of the stimulus category on the number of saccades in 400 to 500 ms was detected; $F(2, 44) = 0.09$ and $P = 0.915$). Furthermore, the TF representation of eye position showed that there was a significant main effect of stimulus position at 4 to 10 Hz horizontal eye movement. This result is similar to a finding by other researchers, of an effect of stimulus position on theta power (3 to 9 Hz), with this effect having a topographical distribution resembling that for eye movements (Carl et al., 2012).

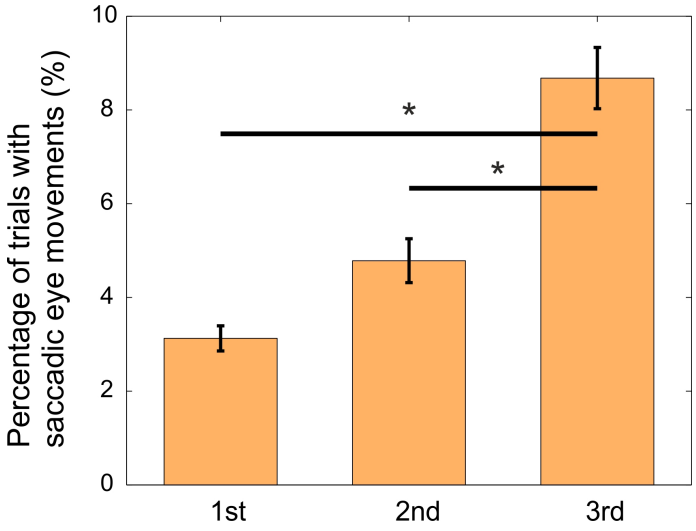

Figure 4.6: The main effect of stimulus position on percentage of saccadic eye movements at 400 to 500 ms during encoding The number of saccadic eye movements increased relative to the working memory load or, in other words, relative to the position of items in the sequence during encoding.

To ensure that the main effect of position in the 400 to 500 ms time-window was not driven by eye movements, I regressed out the part of the power in the TF data which was explained by variations of eye position during the period 400 to 500 ms after stimulus onset. The data used in the regression was the average theta power (4 to 10 Hz) in the channels where the effect was maximal (right temporal channels). Then the main effect of temporal order on the theta power residuals was tested. Residuals are variations in theta power which the variation in the eye position did not explain. After regressing out eye-position related variations, I still found a significant main effect of temporal order on the theta power residuals ($F(2, 44) = 3.8$, and $P = 0.03$). Furthermore, post-hoc t-tests showed that the theta power residuals for the selected channels are significantly different between the last and the middle stimuli in the sequence ($P = 0.24$) but it are not significantly different between the first-position and second-position stimuli ($P = 0.61$) or the first-position and last-position stimuli ($P = 0.12$).

4.4.5 Decoding

One stimulus versus the other two stimuli

I trained classifiers to decode the representation of just the first presentation each stimulus versus a random selection (of equal number) of stimuli from other two types (other). The cross-validated classification performance on the power in the 10 to 45 Hz frequency range indicated that the face was significantly decodable at 110 to 150 ms (cluster level FWE-corrected threshold: $P = 0.001$, peaks at 150 ms; Table 4.1). Decoding the chair was marginally significant at 110 and 130 ms (cluster level FWE-corrected threshold: $P = 0.066$, peak at 130; Table 4.1), and decoding the banana was not significant at any time point (Table 4.1).

The classification of a stimulus vs. others was repeated including power in the 45 to 90 Hz frequency range. The results showed no significant decoding of the face (Table 4.2). The chair was decodable at 330 and 350 ms (cluster level FWE-corrected threshold: $P = 0.045$, peaks at 350 ms; Table 4.2). And decoding the banana was significant at 210 and 230 ms (cluster level FWE-corrected threshold: $P = 0.048$, peaks at 230 ms; Table 4.2) and at 430, 450, and 470 ms (cluster level FWE-corrected threshold: $P = 0.001$, peaks at 450 ms; Table 4.2).

Pairwise decoding

Differences in performance for decoding one stimulus versus others could be caused by large variability in the others group. Therefore, I decoded the representation of each stimulus versus a single other stimulus-type. The results indicated that using 10 to 45 Hz, face vs banana was marginally significantly classifiable at 130 and 150 ms (cluster level FWE-corrected threshold: $P = 0.075$, peaks at 450 ms; Table 4.3), it was significant at 190, 210, and 230 ms (cluster level FWE-corrected threshold: $P = 0.003$, peaks at 210 ms; Table 4.3) and at 410, 430, and 450 ms (cluster level FWE-corrected threshold: $P = 0.003$, peaks at 430 ms; Table 4.3). The face vs. chair was significantly classifiable at 130 to 190 ms (cluster level FWE-corrected threshold: $P < 0.001$, peaks at 150 ms; Table 4.3). And the banana vs chair marginally significant at 250 and 270 ms (cluster level FWE-corrected threshold: $P = 0.082$; Table 4.3).

The classification was again repeated to include power from 45 to 90 Hz. The results indicated that face vs. banana was significantly classifiable at 130, 150 and 170 ms (cluster level FWE-corrected threshold: $P = 0.001$, peak at 150 ms; Table 4.4) and

Table 4.1: Classification performance for decoding representation of a stimulus vs. others using 10 to 45 Hz This table contains the average performance of classifiers across subjects (standard deviations in parenthesis). * is marginally significant and ** is for significant classification - at a cluster level family wise error corrected threshold (0.05).

Time points (ms)	classification performance (%)		
	10 to 45 Hz		
	Face vs Others	Chair vs Others	Banana vs Others
-10	49.02 (1.48)	51.91 (1.36)	48.36 (1.74)
10	51.27 (1.56)	50.88 (1.48)	50.04 (1.83)
30	52.01 (1.66)	48.92 (1.40)	48.42 (1.41)
50	54.20 (1.66)	49.99 (1.50)	46.33 (1.52)
70	51.67 (1.69)	50.07 (1.41)	48.87 (1.44)
90	49.95 (1.61)	52.46 (1.29)	49.74 (1.59)
110	55.18 (1.36)**	50.45 (1.41)*	49.69 (1.85)
130	53.22 (1.53)**	54.33 (1.75)*	50.60 (1.51)
150	56.44 (1.53)**	51.81 (1.77)	51.43 (1.74)
170	51.35 (1.36)	52.59 (1.71)	50.59 (1.59)
190	52.17 (1.48)	54.40 (2.36)	54.15 (2.12)
210	51.40 (1.02)	54.51 (1.91)	51.26 (1.32)
230	52.54 (1.39)	51.80 (1.69)	51.25 (1.51)
250	52.86 (2.28)	49.66 (1.52)	50.79 (1.28)
270	50.29 (1.17)	53.09 (1.34)	53.33 (1.74)
290	50.65 (1.38)	52.28 (1.86)	51.03 (1.56)
310	52.78 (1.92)	52.20 (1.73)	51.53 (1.05)
330	51.87 (1.06)	50.20 (1.95)	49.06 (1.90)
350	50.82 (1.77)	48.81 (1.64)	52.50 (1.57)
370	49.44 (1.54)	50.36 (1.81)	51.52 (1.98)
390	49.99 (1.44)	50.44 (1.68)	50.43 (1.33)
410	50.64 (1.60)	51.10 (1.82)	51.60 (1.84)
430	51.76 (1.55)	52.10 (1.96)	51.20 (1.85)
450	54.65 (1.66)	49.11 (1.88)	54.19 (1.57)
470	50.77 (1.32)	49.24 (1.64)	50.70 (1.90)
490	52.99 (1.96)	52.53 (1.65)	52.18 (1.24)

Table 4.2: Classification performance for decoding representation of a stimulus vs. others using 10 to 90 Hz This table contains the average performance of classifiers across subjects (standard deviations in parenthesis). * is marginally significant and ** is for significant classification - at a cluster level family wise error corrected threshold (0.05).

Time points (ms)	classification performance (%)		
	10 to 90 Hz		
	Face vs Others	Chair vs Others	Banana vs Others
-10	51.23 (2.01)	51.25 (1.50)	49.98 (1.35)
10	52.57 (1.27)	51.71 (1.46)	51.31 (1.53)
30	50.24 (1.67)	51.66 (1.45)	46.79 (1.66)
50	46.08 (1.21)	50.17 (1.53)	51.18 (1.63)
70	48.70 (1.91)	50.63 (1.41)	52.34 (1.48)
90	52.81 (1.74)	48.30 (1.31)	52.41 (1.30)
110	51.55 (1.90)	51.25 (1.47)	47.60 (2.00)
130	51.45 (1.62)	53.14 (1.66)	50.28 (1.61)
150	49.78 (1.55)	53.04 (2.25)	50.17 (1.27)
170	52.48 (1.70)	50.57 (2.25)	51.42 (1.56)
190	51.75 (1.96)	52.50 (1.70)	53.67 (2.14)
210	50.98 (2.07)	53.64 (1.62)	54.16 (1.53)**
230	50.63 (1.49)	50.34 (1.51)	54.11 (1.44)**
250	52.43 (1.37)	51.04 (1.89)	49.30 (1.46)
270	50.61 (1.27)	50.74 (1.75)	49.30 (0.98)
290	49.60 (1.67)	50.72 (1.19)	51.89 (1.42)
310	48.61 (1.91)	52.36 (1.97)	49.78 (1.10)
330	52.59 (2.33)	52.34 (1.32)*	51.18 (1.93)
350	52.40 (2.17)	53.82 (1.25)*	52.32 (2.37)
370	50.21 (1.81)	51.25 (1.41)	51.25 (1.34)
390	50.49 (1.46)	50.93 (1.46)	52.17 (1.60)
410	49.37 (1.54)	50.99 (2.40)	52.03 (1.83)
430	52.82 (1.54)	51.67 (1.20)	55.66 (2.03)**
450	52.36 (1.38)	49.86 (1.85)	54.89 (1.23)**
470	52.45 (2.06)	50.14 (1.67)	53.76 (1.60)**
490	52.85 (1.49)	53.27 (1.47)	49.34 (1.97)

Table 4.3: Classification performance for pairwise decoding using 10 to 45 Hz This table contains the average performance of classifiers across subjects (standard deviations in parenthesis). * is marginally significant and ** is for significant classification - at a cluster level family wise error corrected threshold (0.05).

Time points (ms)	classification performance (%)		
	10 to 45 Hz		
	Face vs Banana	Face vs Chair	Banana vs Chair
-10	53.44 (1.53)	51.17 (1.75)	47.75 (1.61)
10	50.27 (1.28)	51.90 (0.88)	52.71 (1.76)
30	50.26 (1.61)	48.68 (1.71)	47.29 (1.56)
50	47.82 (1.12)	49.16 (1.47)	47.14 (1.29)
70	52.50 (1.16)	49.91 (1.72)	50.28 (1.24)
90	51.90 (1.33)	49.88 (1.20)	49.78 (1.39)
110	52.15 (1.42)	51.62 (1.31)	49.12 (1.06)
130	53.17 (1.51)*	55.15 (1.50)**	50.82 (1.24)
150	55.25 (1.55)*	57.10 (1.49)**	48.89 (1.92)
170	53.17 (1.82)	56.56 (2.44)**	51.04 (2.23)
190	53.35 (1.91)**	54.59 (1.94)**	56.21 (2.50)
210	53.58 (1.73)**	54.91 (1.69)**	53.05 (2.08)
230	52.78 (1.46)**	51.76 (1.01)	50.98 (2.00)
250	51.94 (1.46)	54.27 (1.52)	52.84 (1.31)*
270	50.22 (1.49)	49.94 (1.28)	53.33 (1.55)*
290	52.07 (1.89)	50.71 (1.68)	51.08 (1.59)
310	49.85 (1.16)	51.74 (1.78)	51.71 (1.79)
330	50.49 (1.63)	51.02 (1.82)	53.42 (2.31)
350	50.99 (1.77)	51.94 (1.10)	50.53 (2.06)
370	51.81 (1.42)	52.06 (1.46)	49.17 (1.58)
390	50.57 (1.68)	51.27 (1.73)	52.50 (1.85)
410	53.58 (1.55)**	54.04 (1.70)	52.07 (1.75)
430	54.59 (1.52)**	52.13 (1.75)	49.14 (1.44)
450	53.32 (1.49)**	51.70 (1.37)	49.79 (1.75)
470	52.49 (2.24)	50.01 (1.96)	49.04 (1.65)
490	53.75 (1.62)	54.51 (1.70)	52.27 (1.20)

Table 4.4: Classification performance for pairwise decoding using 10 to 90 Hz This table contains the average performance of classifiers across subjects (standard deviations in parenthesis). * is marginally significant and ** is for significant classification - at a cluster level family wise error corrected threshold (0.05).

Time points (ms)	classification performance (%)		
	10 to 90 Hz		
	Face vs Banana	Face vs Chair	Banana vs Chair
-10	50.99 (1.73)	50.18 (1.79)	50.22 (1.67)
10	52.91 (1.42)	51.69 (1.14)	52.54 (1.09)
30	49.58 (1.91)	48.90 (1.20)	48.26 (0.72)
50	49.72 (1.39)	50.74 (1.72)	51.97 (1.94)
70	49.13 (1.61)	49.96 (1.89)	51.20 (1.61)
90	52.33 (1.56)	49.91 (1.70)	48.57 (1.86)
110	52.15 (1.54)	53.67 (1.58)**	50.48 (2.17)
130	52.02 (1.02)**	56.91 (1.67)**	51.04 (1.27)
150	53.95 (1.69)**	56.64 (1.48)**	50.74 (2.18)
170	54.08 (1.79)**	54.26 (2.05)**	52.39 (2.43)
190	51.38 (2.16)	53.01 (1.38)**	55.04 (2.17)**
210	53.63 (1.93)*	53.50 (1.49)**	54.80 (1.70)**
230	53.19 (1.64)*	47.79 (1.22)	54.13 (1.57)**
250	51.96 (1.63)	52.44 (1.40)	48.55 (1.34)
270	52.04 (1.48)	51.39 (1.14)	53.56 (1.14)*
290	50.69 (1.63)	51.19 (2.00)	54.23 (1.67)*
310	48.19 (1.33)	52.52 (1.55)	52.05 (1.92)
330	52.49 (1.31)	52.40 (1.33)	52.59 (1.87)
350	51.40 (1.57)	52.44 (1.60)	50.02 (1.77)
370	52.06 (1.46)	51.65 (1.17)	51.96 (1.41)
390	48.24 (1.88)	49.85 (1.28)	52.85 (1.64)
410	53.65 (2.10)	50.18 (1.92)	51.32 (1.55)
430	52.82 (1.22)	53.39 (1.80)	53.24 (1.88)
450	52.91 (1.85)	52.11 (1.75)	49.30 (1.70)
470	52.13 (2.25)	50.19 (1.37)	50.74 (1.40)
490	52.92 (2.43)	54.68 (1.09)	52.69 (2.00)

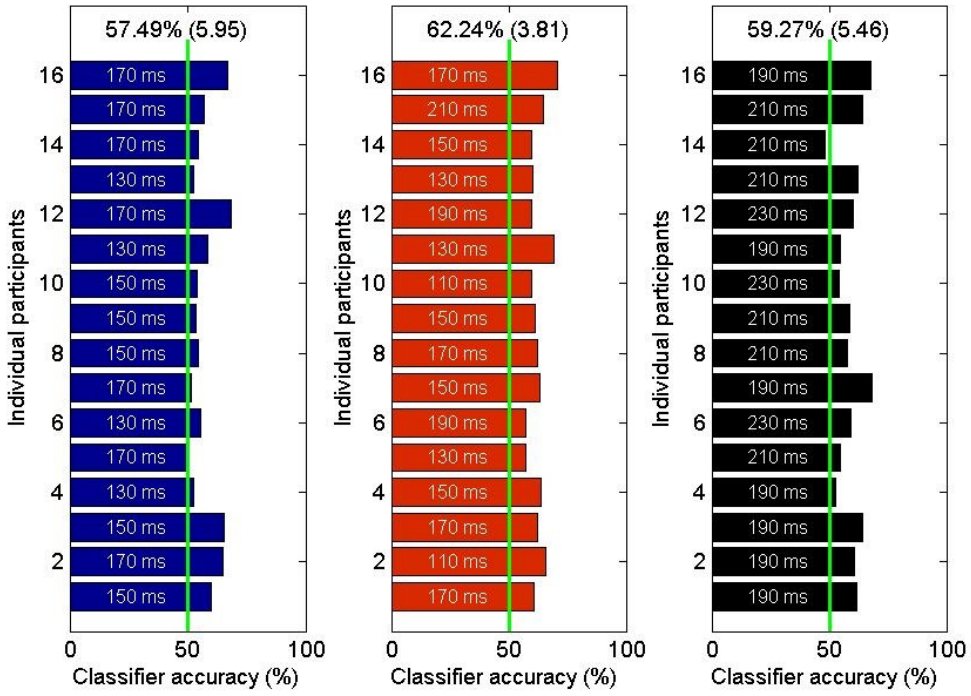

Figure 4.7: Selected decoding accuracies for each participant for decoding each stimulus versus another using the representation in 10 to 90 Hz frequency band. The selection of classifiers was based on maximum performance within the temporal cluster, when the representations of stimuli were decoded. The numbers in white represent the selected time point for each participant. The average performance (with the standard deviation given in parenthesis) is written in black type above the plot. The green line indicates 50 % chance. The plot for decoding face vs banana is shown in blue, face vs chair in red, and banana vs chair in black.

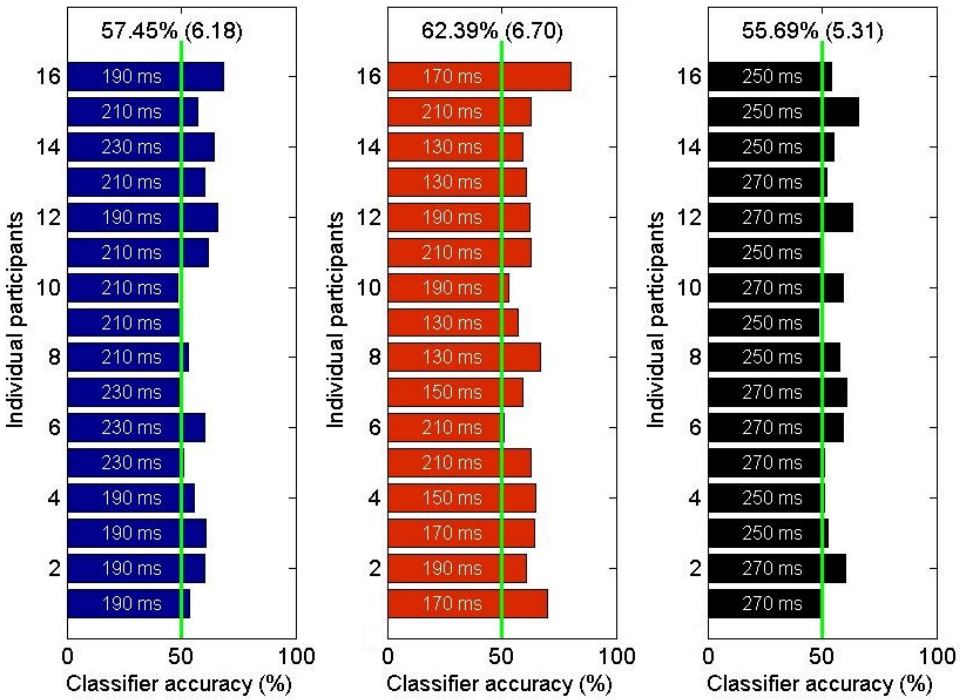

Figure 4.8: Selected decoding accuracies for each participant for decoding each stimulus versus another stimulus using the representation in the 10 to 45 Hz frequency band. The selection was based on maximum performance within the temporal cluster, when the representations of stimuli were decoded. The numbers in white are the selected time point for each participant. The averaged performance (and the standard deviation is in parenthesis) is written in black type above the plot. The green line indicates 50 % chance. The plot for decoding face vs. banana is shown in blue, face vs. chair in red, and banana vs. chair in black.

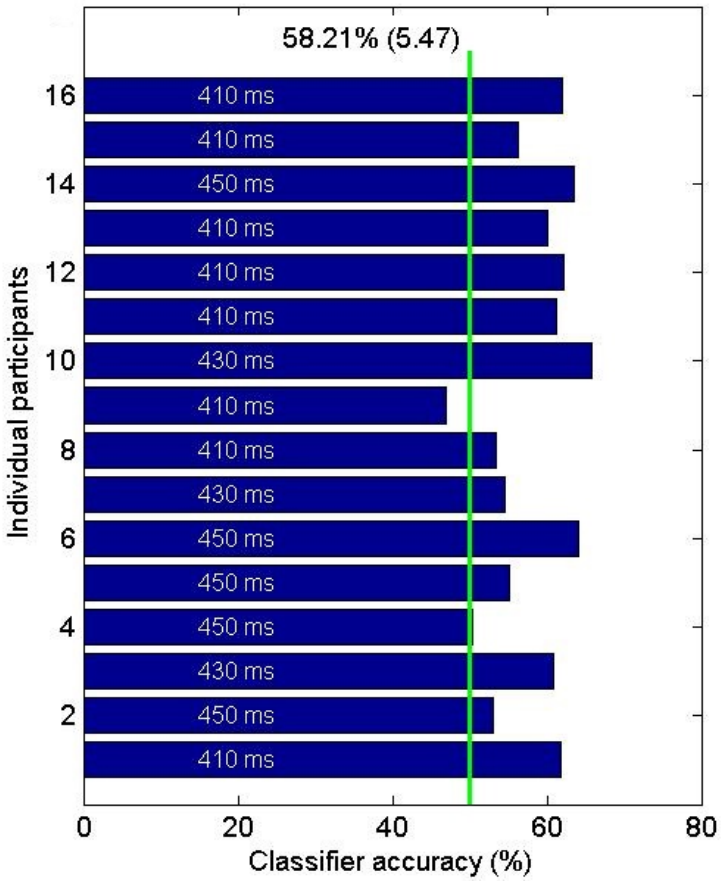

Figure 4.9: Selected face vs. banana decoding accuracy for each participant using the representation in the 10 to 45 Hz frequency band. The selection was based on maximum performance within 410 to 450 ms temporal cluster, when the representations of stimuli were decoded. The numbers in white are the selected time point for each participant. The averaged performance (and the standard deviation is in parenthesis) is written in black on top of the plot. The green line indicates 50 % chance.

marginally significant at 210 and 230 ms (cluster level FWE-corrected threshold: $P = 0.056$; Table 4.4). The face vs. chair was decoded significantly at 110 to 210 ms (cluster level FWE-corrected threshold: $P < 0.001$; Table 4.4). And the banana vs. chair was decoded significantly at 190, 210, and 230 ms (cluster level FWE-corrected threshold: $P = 0.001$; Table 4.4), and marginally significantly decoded at 270 and 290 ms (cluster level FWE-corrected threshold: $P = 0.050$; Table 4.4).

Finally, for each participant I selected the classifier with maximum performance within a time window where decoding was frequently significant - about 100 to 300 ms. As a result, the average selected decoding using 10 to 90 Hz for face vs banana was 57.49% (SD = 5.46), for face vs chair was 62.24% (SD = 3.81), and for banana vs chair was 57.49%(SD = 5.95) (Figure 4.7). The same analysis using 10 to 45 Hz resulted in 57.45% (SD = 6.18) for face vs banana, 62.39% (SD = 6.70) for face vs chair, and 55.69% (SD = 5.31) for banana vs chair (Figure 4.8). Furthermore, the analysis of face vs banana classification, using 10 to 45 Hz in 400 to 450 ms, resulted in an average accuracy of 58.21% (5.47) (Figure 4.9). Interestingly, this is the same time window when there was a main effect of item position in the sequence.

4.5 Discussion

In this chapter I successfully detected the representation of stimuli at 100 to 200 ms using univariate ANOVA and multivariate pattern classifiers (based on power from 10 to 90 Hz). The significant main effect of the type of stimulus on ERFs was detected over occipital and temporal channels which is coherent with previous research on categorically specific representations of visual stimuli (J. Liu et al., 2002; Gao et al., 2013; Carlson et al., 2013). This categorical representation occurs only in the

the 100 to 200 ms time window (chapter 3 and Carlson et al., 2013).

Studies on familiarity (Curran, Tanaka, & Weiskopf, 2002) and semantic memory (Dien, Michelson, & Franklin, 2010) suggest that semantic representations of familiar stimuli emerge, or are retrieved, at about 400ms (Kutas & Federmeier, 2011). It is therefore possibly that in my experiment the MVPC is decoding semantic representations of face and banana at 400 ms. In this time window the decoding including chair (face vs. chair or banana vs. chair) was not significant. In addition, behavioural measures showed that participants' working memory abilities for the chair stimulus were significantly below those for the face and banana, possibly because images of the face and banana from multiple points of view were less repetitive and similar than the images of the chair from different perspectives. Such similarity might have led to stronger repetition suppression for chair images than for face and banana images, and might thus have reduced the decodability of the chair representation.

At 400 to 500 ms the ERFs over the frontal channel showed a main effect of stimulus position in the sequence. This effect relates to the number of items having been presented in the encoding period, and it can therefore be considered an effect of working memory load, similar, for example, to the P300 (Morgan et al., 2008). Previous studies that have used arrays of visual items (where all items are presented at the same time, and for a very short duration of, for example, 100 ms) have shown that the ERP elicited during the delay reflects the number of items in working memory; stronger deflections reflect maintenance of larger numbers of items (Vogel & Machizawa, 2004). Specifically, associations between working memory load and the P300 ERP response (McEvoy, Smith, & Gevins, 1998) and slow wave oscillations (Gevins et al., 1998) during perception have been reported. Furthermore, frontal theta power has been shown to increase with memory load (Jensen & Tesche, 2002).

There is a consistency between these findings and those in my own experiment, where the theta power between 400 and 500 ms during presentation of the last stimulus in the sequence was significantly higher than the during presentation of the stimuli occurring in the first or second position in the sequence.

The working memory load effect can also be interpreted as a representation of temporal order (akin a relative-temporal context). Based on the temporal context model (TCM), the representation of an event comprises a combination of the current representation and the representation of temporal information/context. Indeed temporal context is defined at least in part by the sequence of preceding events during an episode; thus based on definition, temporal context is the collection of events that happened in the adjacent times (Howard & Kahana, 2002).

In my study, the 400 to 450 ms classification test was based on the representation of the first item in the sequence, thus the working memory load was 1. The remaining questions for future work are: (1) is the representation at 400 to 450 ms representative of item specific (semantic) information and is this irrespective of the temporal position of an item in the sequence? (2) is the representation reflective of working memory load? (3) are the 400 to 450 ms representations replayed during the maintenance delay period?

To address the first of those questions, I would, in a follow up study, carry out a cross-validation of the classification test using the representations of stimuli presented in second and third position in the sequence. Specifically, the classifier trained on the first epoch will be used to test the representation of presented items during other epochs. If decoding is successful, it would demonstrate that there is item specific information represented during those epochs. For the second question, suc-

cessful decoding of individual items would also show that the representations were not dependent on temporal information. If the test reveals representation of multiple items corresponding to the content of what is being held in working memory, and to the size of the working memory load, during that time-period, then it could be argued that, at 400 to 500 ms after onset of each stimulus, the working memory load system accumulates representations of items being held in working memory. And for the final question, classifiers trained on the representations detected at the early and later time windows would be used to investigate the replay of contents of working memory during maintenance.

Chapter 5

Population level inference for multivariate MEG analysis

5.1 Precis

Multivariate analysis is a powerful technique for analysing MEG data. An outstanding issue, however, is the problem of group-level inference for multivariate MEG analysis. Here a solution is proposed based on model-scoring at the subject-level and random effects Bayesian model selection at the group-level. I have applied this approach in a Canonical Variates Analysis (CVA) of beamformer reconstructed MEG data in source space. The MEG data is from the associative-recognition experiment in Chapter 3. I used the epoch of the encoding phase, when images of faces were being presented. I decoded the oscillatory activities in primary visual cortex (V1) and fusiform gyrus (FFG) into pre-stimuli and post-stimuli activities.

Here CVA estimates the multivariate patterns of activation, in the frequency domain, that correlated most highly with the experimental design. The number of significant canonical vectors determined the order of a CVA model, and model com-

parison provided machinery for inferring the optimal order at the group level. In this chapter, I demonstrate this approach by testing how the use of different feature sets affects decoding of experimental conditions, and identifying those feature sets that were maximally discriminative of the conditions (pre- or post- stimuli). [1]

5.2 Introduction

Multivariate group inference has been used to study oscillatory representations of stimuli, for example, visual stimuli at the time of perception (Newman & Norman, 2010; Duncan et al., 2010; Carlson et al., 2013) or the replay of oscillatory patterns during memory tasks (Fuentemilla et al., 2010; Jafarpour et al., 2013). A major problem with such multivariate analyses is the identification of discriminative data features from a high dimensional measurement space. In the above studies, the set of discriminatory features was allowed to vary from subject to subject. This between-subject variability, however, makes it difficult to interpret experimental findings in terms of a consistent set of underlying cognitive processes.

In this chapter I presented a principled approach to systematically select the most discriminatory and minimally complex feature set that is consistent over subjects i.e. at the 'group level'. This analysis therefore enables systematic inference on the dimensions of feature-spaces. (If there is no dependence between data features and experimental conditions the inferred dimension will be zero.)

This framework is based on CVA that models multivariate dependencies between a set of class labels and data features. The order of the CVA model is then based

[1] Jafarpour A, Barnes G, Fuentemilla L, Duzel E, Penny WD (2013) "Population Level Inference for Multivariate MEG Analysis". *PLoS ONE* 8(8): e71305. doi:10.1371/journal.pone.0071305

on the number of significant canonical vectors, as determined by the Bayesian Information Criterion (BIC) (Zhou, 2008; Chalise & Fridley, 2012). Absence of a multivariate dependence (or no significant decodability) is indicated by the zeroth order model (null model or model 0) being the most likely. Here I applied CVA to beamformer reconstructed MEG data in source space (G. Barnes et al., 2011). Nevertheless, in principle, it could be applied to data in sensor space or to data after various transformations, including the use of principal or independent component analysis (Onton & Makeig, 2006).

The model ranking approach allows us to test, at the group level, both whether there is a multivariate dependence between data features and experimental condition, and if there is, to find which feature sets maximise the strength of this dependence.

To test the consistency of these multivariate dependencies over a group of subjects, I used random effects Bayesian model selection (Stephan, Penny, Daunizeau, Moran, & Friston, 2009), based on the BIC values. I illustrated this method to determine the spectral resolution (number of frequency bands) that maximizes decodability of data features into the experimental conditions. I used MEG power spectra at each voxel in source space, within the regions of interest (ROIs), the primary visual cortex (V1) and fusiform gyrus (FFG), and the experimental conditions indicating whether the data were from a pre- or post-stimulus epoch of a simple visual processing paradigm.

5.3 Methods and materials

This section describes the proposed data processing pipeline, comprised of five steps:

1. **MEG Source Reconstruction,** In this chapter the features of the MEG signal are power spectra at the source level. But in general the method can also be applied in sensor space, or more generally, it can be any feature of the MEG data, such as phase, amplitude and/or nonlinear measures.

2. **Canonical Variates Analysis,** A CVA model at each point in source space was applied in order to generate a brain mapping. The maps indicate which areas show consistent relationships between multivariate data features and experimental condition.

3. **Bayes factors,** The order of a CVA model is determined by the number of canonical vectors. This step computes the evidence of a model with m canonical vectors in relation to the evidence of a model with zero canonical vectors. The ratio of these evidences is known as a Bayes factor.

4. **Feature Set Selection,** The optimal model will depend not only on the number of canonical vectors but also on the features to which these vectors map. Here I compared models with single canonical vectors but with a different fractionation of the MEG power spectrum.

5. **Random Effects Bayesian Model Selection (RFX-BMS),** The previous steps are applied to produce Bayes factor maps for each subject and model. The maps allow for single alternative models to be compared with a null model, or for any number of models to be simultaneously compared with each other. This final step computes the frequency with which models are used in the population.

The following subsections describe each of the above steps in more detail and then the data analysis process which demonstrates the proposed approach is explained.

5.3.1 MEG Source Reconstruction

I source reconstructed data for each subject using the SPM8 implementation of the Linearly Constrained Minimum Variance (LCMV) beamformer (Robinson & Vrba, 1999; G. R. Barnes & Hillebrand, 2003; Sekihara, Nagarajan, Poeppel, Marantz, & Miyashita, 2002). The software for source reconstruction, and computation of Bayes factors for CVA models is available in the SPM Beamforming toolbox (`http://code.google.com/p/spm-beamforming-toolbox/`). This produces a log Bayes factor image for each subject and model. The software for implementing Random effects Bayesian model selection is available in the latest release of SPM (W. Penny & Flandin, 2005) (`http://www.fil.ion.ucl.ac.uk/spm/software/`). This takes the log Bayes factor images for all subjects and produces expected frequency maps (`*_xpm.img`) and, optionally, exceedance probability maps (`*_epm.img`).

The forward model used in source reconstruction was defined using an inverse normalized canonical head-shape brain for all subjects (Litvak et al., 2011). At each source location we selected the orientation that maximises projected power (Sekihara et al., 2002) which gives a single weight vector for each source location. Briefly, the weights for location s were given by

$$w_s = (L_s C^{-1} L_s^T)^{-1} L_s C^{-1} \tag{5.1}$$

where $L_s \in \mathbb{R}^{1\times m}$ is the lead field matrix for m channels at source location s and $C \in \mathbb{R}^{m\times m}$ is the sensor covariance matrix. This corresponds to an LCMV beamformer with zero for the regularisation parameter (G. R. Barnes & Hillebrand, 2003).

Accordingly, the source level estimate of activity for trial n at location s is given by

$$y = w_s B_n \tag{5.2}$$

where $B_n \in \mathbb{R}^{m \times p}$ comprises p complex valued Fourier coefficients describing the signal at m MEG sensors on trial $n \in 1, ..., N$. In the next section we go on to look at multivariate dependence between the experimental design and the spectral features and this source level estimate across the brain.

5.3.2 Canonical Variates Analysis

CVA is a method for detecting dependencies between a set of variables $X \in \mathbb{R}^{N \times q}$, and a set of variables $Y \in \mathbb{R}^{N \times v}$. The aim of CVA is to find the linear projections of X and Y with maximal correlation. Given $z_x = XU_x$ and $z_y = YU_y$, one can compute the canonical correlation

$$r = \frac{\|z_x^T z_y\|^2}{\|z_x^T z_x\| \|z_y^T z_y\|}. \tag{5.3}$$

The projections U_x and U_y which maximise this correlation are known as the canonical vectors and the resulting z_x and z_y are the canonical variates. If

$$\Sigma = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix} \tag{5.4}$$

is the sample covariance matrix and V_Y and V_X are the left and right singular vectors of

$$\Sigma_{XX}^{1/2} \Sigma_{XY} \Sigma_{YY}^{1/2} \tag{5.5}$$

in decreasing order, then the canonical vectors can be computed as $U_Y = \Sigma_{YY}^{-1/2} V_Y$ and $U_X = \Sigma_{XX}^{-1/2} V_X$ (Chatfield & Collins, 1980). There are $i = 1..h$ pairs of canonical vectors where $h = \min q, v$. The canonical correlations r_i for $i = 1..h$ are used to compute Bayes factors, as described in the following section.

5.3.3 Bayes Factors

The dimension of a CVA model is given by the number of significantly non-zero canonical vectors. If a linear multivariate dependence between X and Y exists, then the dimension of the corresponding CVA model is non-zero. Thus one can test for linear multivariate dependence by estimating CVA model dimension.

A standard approach from classical inference here is Bartlett's test for dimensionality (Chatfield & Collins, 1980). However, to our knowledge, there is no simple way to carry over these results to the group level. I therefore used a Bayesian method, as this integrates seamlessly with established methods for group level inference (see final subsection).

These Bayesian methods first compute the evidence for a model, $p(Y|m)$, with m canonical vectors. Various methods exist for computing the Bayesian model evidence for a CVA model. These include the Bayesian Information Criterion (BIC) (Zhou, 2008; Chalise & Fridley, 2012) and variational approximations (C. Wang, 2007). In this project I used BIC approximation.

Bayesian Information Criterion formulation

The BIC value is calculated as following. If there is no relation between dependent variable (or 'data') Y and independent variable X, then the log-likelihood of the data is

$$\log p(Y) = -\frac{N}{2} \log |\Sigma_{YY}| \tag{5.6}$$

where Σ_{YY} is the data covariance. If there is a relation between X and Y then the log-likelihood can be calculated as follows. The maximum likelihood coefficients are $\beta_{ML} = (X^T X)^{-1} X^T Y$ and the log-likelihood is

$$\log p(Y|\beta_{ML}) = -\frac{N}{2} \log |\Sigma_{Y|X}| \tag{5.7}$$

where $\Sigma_{Y|X} = \Sigma_{YY} - \Sigma_{XY}^T \Sigma_{XX}^{-1} \Sigma_{XY}$, Σ_{XY} is the covariance between X and Y, and Σ_{XX} is the covariance of X. The log-likelihood ratio, Λ, is therefore

$$\begin{aligned} \Lambda &= \log \frac{p(Y|\beta_{ML})}{p(Y)} \\ &= \frac{N}{2} \log |\Sigma_{Y|X}^{-1} \Sigma_{YY}|. \end{aligned} \tag{5.8}$$

If Σ_i is the i-th eigenvalue of $\Sigma_{Y|X}^{-1} \Sigma_{YY}$ then

$$\Lambda = \frac{N}{2} \sum_{i=1}^{h} \log \sigma_i \tag{5.9}$$

where $h = \min(q, v)$. This is also known as Wilk's Lambda (Chatfield & Collins, 1980). Also given the quantity

$$\Lambda_{j,t} = \frac{N}{2} \sum_{i=j}^{t} \log \sigma_i \tag{5.10}$$

, $\Lambda_{1,t}$ is the log-likelihood ratio for a CVA model with t canonical variates. The quantity $\Lambda_{j,t}$ was used to compute the BIC and can also be expressed in other forms. Later in this chapter I showed how BIC was computed in an implementation, and finally showed how it was related to canonical correlations.

A second expression for $\Lambda_{j,t}$ can be derived as follows. Let $\Sigma_{YY} = \Sigma_{\hat{Y}\hat{Y}} + \Sigma_{Y|X}$, where $\Sigma_{\hat{Y}\hat{Y}}$ is the covariance explained by the model and $\Sigma_{Y|X}$ is the covariance not explained by the model. Then if λ_i are eigenvalues of $\Sigma_{Y|X}^{-1}\Sigma_{\hat{Y}\hat{Y}}$ then the above relationship can be used to show that $\sigma_i = \lambda_i + 1$ (see Appendix B of Friston, Ashburner, Kiebel, Nichols, & Penny, 2007). Hence an alternative expression for Wilk's Lambda is

$$\Lambda = \frac{N}{2} \sum_{i=1}^{h} \log(1 + \lambda_i) \tag{5.11}$$

This expression was used in the current study and it has been implemented in the SPM software (Friston et al., 2007).

Accordingly, $\Sigma_{\hat{Y}\hat{Y}}$ can be formed directly from model predictions

$$\begin{aligned} \hat{Y} &= X\beta_{ML} \\ \Sigma_{\hat{Y}\hat{Y}} &= \hat{Y}^T\hat{Y} \end{aligned} \tag{5.12}$$

and $\Sigma_{Y|X}$ from the residuals

$$\begin{aligned} R &= Y - \hat{Y} \\ \Sigma_{Y|X} &= R^T R. \end{aligned} \tag{5.13}$$

Relation to Canonical Correlation

The i-th canonical correlation can be expressed as $r_i = \sqrt{\frac{\lambda_i}{\lambda_i+1}}$. Hence, a third equivalent form for the log likelihood ratio is

$$\Lambda = -\frac{N}{2}\sum_{i=1}^{h}\log(1-r_i^2). \tag{5.14}$$

In summary, one can write

$$\begin{aligned}\Lambda_{j,t} &= \frac{N}{2}\sum_{i=j}^{t}\log\sigma_i \\ &= \frac{N}{2}\sum_{i=j}^{t}\log(1+\lambda_i) \\ &= -\frac{N}{2}\sum_{i=j}^{t}\log(1-r_i^2).\end{aligned} \tag{5.15}$$

This last expression appeared in (Zhou, 2008; Chalise & Fridley, 2012).

The log evidence for a model with no parameters (null model) is simply the log likelihood of the data, $L_0 = \log p(Y)$. The log evidence for model m with parameters β is given by $L_m = \log\int p(Y|\beta)p(\beta)d\beta$. This can be approximated by the Bayesian Information Criterion (BIC) as

$$BIC = \log p(Y|\beta_{ML}) - \frac{k}{2}\log N \tag{5.16}$$

where k is the number of parameters in the model and β_{ML} are the maximum likelihood parameters. A Bayes factor is the ratio of model evidences. Here Bayes factor is defined as

$$BF(m) = \frac{p(Y|m)}{p(Y)}. \tag{5.17}$$

Log Bayes factors is, therefore, approximated as differences in the BIC scores. Under BIC, the log Bayes factor for a CVA model of dimension m versus a model with dimension zero (null model) is given as

$$LogBF(m)_{BIC} = \Lambda_{1,m} - \frac{k}{2}\log N, \tag{5.18}$$

where

$$\Lambda_{1,m} = -\frac{N}{2}\sum_{i=1}^{m}\log(1 - r_i^2), \tag{5.19}$$

and N is the number of data points and r_i are the canonical correlations at each dimension i . This expression has been used in previous studies (Zhou, 2008; Chalise & Fridley, 2012) and is derived from equation 5.15. The estimated model order is the one which has the largest LogBF. Negative values of $LogBF(m)$ express evidence in favour of the null model. Intuitively, better CVA models will have stronger canonical correlations (r_i) and fewer parameters (k).

5.3.4 Feature Set Selection

It is also possible to compute Bayes factors for models that have the same number of canonical vectors but are supplied with different feature sets. Bayesian model comparison here allows the models to vary but the data must stay the same. Feature set selection therefore requires that I set up CVA models such that Y is a design matrix encoding experimental conditions and X are independent variables comprising the neuroimaging data features (in other words the traditional roles of these variables (X and Y) are switched to make it clear that I am searching for optimal data features for a fixed experimental design). In this paper these features are d-dimensional power spectra. I then compute $LogBF(d)$ images where each is the log Bayes Factor for a model with a single canonical vector and d-dimensional features X, versus a model with zero canonical vectors. One can then use the same Bayes factor images

to compare different feature dimensions. For example, for pairwise comparisons

$$\begin{aligned} BF(d_1, d_2) &= \frac{p(Y|d_1)}{p(Y|d_2)} \\ &= \frac{p(Y|d_1)}{p(Y)} \frac{p(Y)}{p(Y|d_2)} \end{aligned} \tag{5.20}$$

Hence

$$LogBF(d_1, d_2) = LogBF(d_1) - LogBF(d_2) \tag{5.21}$$

I can also implement multi-way comparisons as described in the next section.

5.3.5 Random effects Bayesian model selection

Random effects Bayesian model selection (RFX-BMS) (Stephan et al., 2009) views the assignment of models to subjects as a random process in which each subject is assigned to model i with probability f_i. Here f_i is the frequency with which model i is used in the population from which the subjects were drawn.

The Bayesian algorithm for estimating model frequencies f_i from the table of log model evidence values (Stephan et al., 2009) uses a Dirichlet prior

$$p(f) = \mathsf{Dir}(f; \alpha^0) \tag{5.22}$$

with 'count parameters' $\alpha_m^0 = 1$. These parameters can be thought of as corresponding to the assumption of having previously observed one instance of each model type. These parameters produce a flat prior. The posterior is approximated to also be a Dirichlet

$$p(f|\tilde{Y}) = \mathsf{Dir}(f; \alpha) \tag{5.23}$$

where $\tilde{Y}$ indicates data from all subjects. The count parameters α_m, are initialised as α_m^0, and then updated iteratively as follows

$$\begin{aligned} u_{nm} &= \exp\left[\log p(y_n|m) + \psi(\alpha_m) - \sum_m \psi(\alpha_m)\right] \\ g_{nm} &= \frac{u_{nm}}{\sum_{m'} u_{nm'}} \\ \alpha_m &= \alpha_m^0 + \sum_n g_{nm} \end{aligned} \tag{5.24}$$

where $\psi()$ is the digamma function (Press, Flannery, Teukolsky, & Vetterling, 1991). Here $\log p(y_n|m)$ is the entry in the log evidence table from the n-th subject (row) and m-th model (column). The quantity g_{nm} is the posterior probability that subject n used the m-th model. This is 'posterior' as in after seeing the model evidence table and, implicitly, the data from all subjects $\tilde{Y}$.

This algorithm can also be applied to a table of log Bayes factor values, as long as the Bayes factors have all been computed with respect to the same common model. In this paper, the common model is the null CVA model with zero canonical variates.

The goal of RFX-BMS is to estimate f_i using a table of model evidence scores, or Bayes factors with respect to a common model, from S subjects and K models. Intuitively, if the scores favour model i in 9 out of 10 subjects, then f_i will be estimated to be about 0.9. However, the estimate of f_i is also influenced by the degree to which models are favoured. For example, if for the 10th subject the score is greatly in favour of a different model then the estimate of f_i will be commensurately reduced. Given data Y (in practice, a table of logBF values), a posterior distribution, $p(f_i|Y)$, can be estimated using the algorithm described in (Stephan et al., 2009). The mean of this distribution, $< f_i|Y >$ provides an estimate of the model frequencies. This is also referred to as the 'expected frequency'.

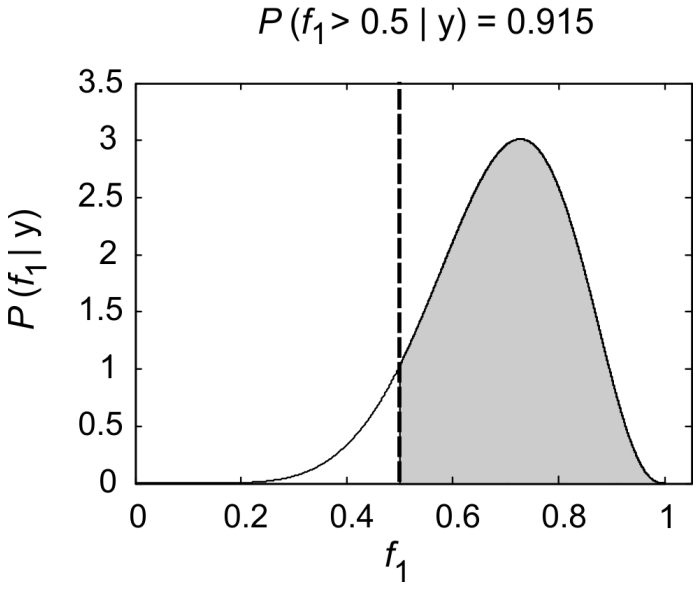

Figure 5.1: This graphic demonstrates random effects model selection in the case of comparing two models. These models have frequencies f_1 and f_2. These frequencies refer to the population from which the subjects were drawn. The figure plots the posterior probability of f_1. The mean of this density is $< f_1|Y >= 0.75$, indicating that 75% subjects use model 1. For two models, the exceedance probability $p(f_1 > f_2|Y) = p(f_1 > 0.5|Y)$ is given by the posterior mass in excess of $f_1 = 0.5$. Here, the exceedence probability is 0.915.

It is also possible to compute the probability that one model frequency exceeds another. For example, when comparing just two models I can compute $\phi_1 = p(f_1 > f_2|Y)$. This is known as the exceedance probability for model 1 over model 2. Figure 5.1 illustrates the concept of an exceedance probability. If one has maps of model evidence over anatomical space, and for multiple subjects, it is possible to produce maps of expected frequencies or exceedance probabilities. In a previous work, for example, Rosa, Bestmann, Harrison, and Penny (2010) plotted Exceedance Probability Maps (EPMs) for univariate General Linear Models fitted to functional MRI data. In this chapter I plotted expected frequency maps for CVA models in MEG source space.

5.3.6 Experimental MEG data analysis

At each point in source space I generated a d-dimensional feature vector of power in d frequency bands. Here I used 6 different features. The frequency bands are as defined in Table 5.1. The average power in the frequency band was computed separately across 1 second before (-1000 to 0) and 1 second after (0 to 1000ms) onset of visual stimulus epochs. I analysed data in two regions of interest, V1 and FFG, defined using the MNI grey matter masks shown in Figure 5.2A. I excluded any voxels which overlapped in the low-resolution source localization grid space (10 millimetres resolution). The FFG mask included 1600 voxels and the V1 mask included 574 voxels. In the proposed formulation Y variable contains class labels with a scalar $+1$ indicating post-stimulus, and a -1 indicating pre-stimulus. I have $v = 1$. Therefore the CVA model has at most a single canonical component, $h = 1$. The X variable contains the d-dimensional power spectra. Thus, the number of parameters in the CVA model is $k = d + 1$.

Table 5.1: Definition of feature space This describes the fractionation of the power spectrum into d separate bands, $d = 1, 3, 5, 7, 9$ and 11.

d	Frequencies(Hz)
1	3-90
3	3-10, 10-30, 30-90
5	3-8, 8-12, 12-30, 30-50, 50-90
7	3-5, 5-8, 8-12, 12-20, 20-30, 30-50, 50-90
9	3-5, 5-8, 8-10, 10-12, 12-20, 20-30, 30-40, 40-50, 50-90
11	3-5, 5-8, 8-10, 10-12, 12-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-90

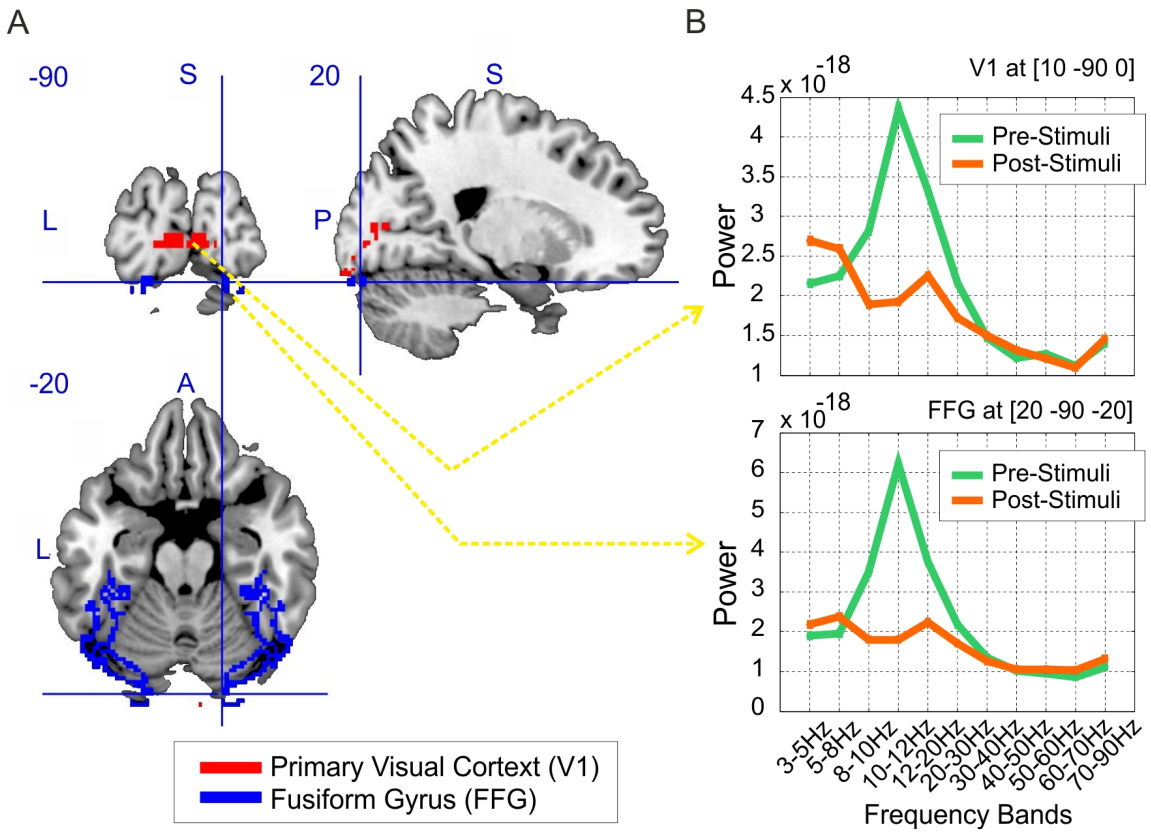

Figure 5.2: Regional Activity (A) the MNI mask for region of interest: grey matter in the primary visual cortex (V1) and fusiform gyrus (FFG), the view is from [20 -90 -20]mm. (B) The power spectrum for pre- and post- stimulus activity across 11 frequency bands for an individual participant. The top plot is the signal from a V1 voxel, [10 -90 0]mm (Talairach coordinates), and the bottom plot is the signal from a FFG voxel, [20 -90 -20]mm.

The matrix Y was prepared in the following way. Let each j index Fourier bins within one of the pre-defined spectral bands ranging from F^{lower} to F^{upper} frequency. Then at each source location s the activity at trial n is formulated as $y_n = \sum_{j=F^{lower}}^{F^{upper}} (w_s B_{nj})(w_s B_{nj})^*$ based on equation 5.3.1 (see Methods Chapter), where $*$ signifies the complex conjugate. In order to give equal weighting to all frequency bands (some of which have markedly less power) for each band I removed the mean value (across all conditions) and normalized the variance (in power) to unity.

For each subject I computed the $logBF(d)$ maps for $d = 1, 3, 5, 7, 9$ and 11. Each $logBF(d)$ map is the log-evidence for a model with a single canonical vector, and d signal features, minus the log-evidence of the null model (no canonical vectors).

5.4 Results

I studied 3 to 90 Hz oscillatory activity in pre- and post- visual stimuli presentation (an unfamiliar face) in source space, specifically in primary visual cortex (V1) and fusiform gyrus (FFG). The aim of this analysis was to decode the differences in pre versus post stimulus activity, based on different feature sets (see Table 5.1 in the Methods and Material section). The hypothesis was that there will be differences elicited by onset of the stimuli in the regions of interest (ROIs). These regions were defined using the anatomical masks based on the MNI brain (Figure 5.2A). In the first step, the MEG data for each individual participant was source localized and the features were defined as average power within the specified frequency bands. Figure 5.2B illustrates the average power across the 11 frequency bands (Model 11) in selected voxels from V1 and FFG.

Table 5.2: Model with different number of features versus the null model Total (and percentage) of voxels favouring higher (than zeroth) order models in more than 90 present of the population using different feature spaces in each of the ROIs.

ROI	Model 1	Model 3	Model 11
FFG (1691 voxels)	64 (8.74%)	615 (84.01%)	103 (14.07%)
V1 (732 voxels)	32 (4.37%)	510 (69.67%)	58 (7.92%)

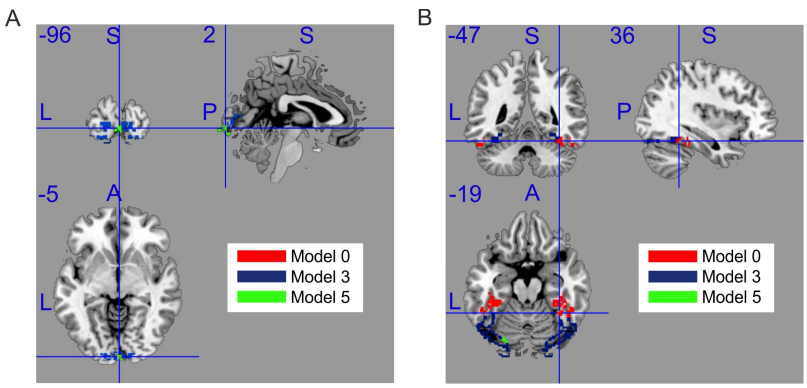

Figure 5.3: Model 3 versus Model 0 exceedance probability map Multivariate maps for V1 and FFG, showing voxels where pre- versus post-stimulus activity are best discriminated by a multivariate (model 3) as opposed to the null (model 0). The color indicates the percentage of population showing the effect in the voxel, exceedance probability. (A) FFG view from [46 -54 -20]mm and color bar is from 0% to 100% (white). (B) V1 view from [6 -81 7]mm (Talairach coordinates) and color bar is from 70% to 100% (white).

5.4.1 Model-d versus model 0

The group level analysis was carried out separately for each of the three feature spaces $d = 1, 3$, and 11. The effect of visual stimulation in the ROI was compared to the null model. Table 5.2 summarized the number of voxels in each ROI which showed a correlation between the experimental design and the feature space in more than 90% of the population. In other words, the percentage of voxels in the ROIs in which the signal can decode the experimental condition. The results indicate that, for instance, in 69% of the voxels within V1 and in 84% of the voxels within FFG Model 3 was preferred to Model 0 across more than 90% of the population. As an example Figure 5.3 shows the expected probability map of model 3 being more frequent than model 0 in the ROIs.

5.4.2 Between models comparison

In the next step I studied which feature space best decoded the signals in the ROI based on the experimental design. In this case I looked for the most likely non-null model across the population by applying the random effect Bayesian model selection procedure to the table of log Bayes factor values from each voxel within ROIs with three columns; one for $LogBF(1)$, one for $LogBF(3)$ and the other for $LogBF(11)$. The frequency of model selection was then averaged over ROI voxels. Figure 5.4 shows both univariate and complex (11 features) multivariate parcellations are relatively poor models to describe the spectral power change. In both V1 and FFG, the univariate parcellation is a poor model because it has a low correlation with the design (although few features); whereas the complex model (11 features) is poor because the improvement in the canonical correlation is not justified by the large number of features (Equation 5.18).

5.5 Discussion

In this Chapter I have proposed a solution to the problem of group-level inference from multivariate modelling of MEG data. The combination of scoring CVA models using BIC, and assessing consistency across a group using Bayesian random effects model inference, provides a principled and computationally efficient solution. I applied this approach on power spectra estimates in V1 and FFG to decode pre- and post-stimuli presentation, using three different feature sets. I was able to show that although all feature-sets provided some degree of discrimination between experimental conditions; the optimal feature set for this population consisted of three approximately classical bands (3-10, 10-30, 30-90 Hz). Although these features sets were selected principally as examples, it is clear that the multivariate models outperformed univariate models in this simple decoding task.

This data analysis focussed on inferring the optimal feature dimension across three exemplar feature sets. But there are also other ways in which random effects Bayesian model selection can be used for finding the optimal feature set. Those include, for instance, fixing the number of features, but changing the feature set (eg. by breaking up spectra in a different way, or using phase/amplitude or more exotic nonlinear features). Or whereas in my analysis I have focussed on regions of interest, one could also study the potential effect of experimental conditions on different brain regions at different times. In this case, the proposed approach can be used across group of participants in combination with a searchlight algorithm (Kriegeskorte, Goebel, & Bandettini, 2006), to find the most relevant spatial and temporal activity in group level.

Random effect Bayesian model selection is beneficial in two important aspects, computational efficiency and the possibility of model comparison. This is because the sig-

nificance of the accuracy does not need to be established using permutation testing. An alternative to the proposed scheme, which does however require cross-validation to compute the accuracies, is mixed effects modelling (Brodersen et al., 2012). This approach furthermore allows comparison between different feature spaces and identifying the most evident feature space over a group. It is concerned only as to whether, for example, more subjects use model A than model B. This does not require that the parameters of the winning model are consistent over that group. For example, here the model comparisons showed multiple voxels in which post-stimulus activity was better discriminated from pre-stimulus activity when the spectrum was described using a triplet (power in low, medium and high frequencies) rather than a scalar (power across all frequencies). This does not necessarily mean that the pattern of frequency responses was consistent over subjects. For example, half the subjects may have increases in low frequency power post-stimulus, and the other half decreases. One way to directly test for this scenario using the same scheme would be to see if a model using a fixed canonical vector over subjects (essentially a univariate test with data projected onto a single canonical vector) has more evidence than a model in which the canonical vector is allowed to vary (as here). Similar model comparison approaches could be used to test for differences (e.g. in the feature set) between different study groups, such as patients and control groups.

A limitation of this approach is that it requires a good approximation of the Bayesian model evidence, at the subject level. Whilst this is straightforward for linear models (Bishop, 2006a) it is more problematic for nonlinear models. Approaches based on Laplace approximations do, however, provide good approximations if the number of data points is large with respect to the number of model parameters (W. D. Penny & Roberts, 1998).

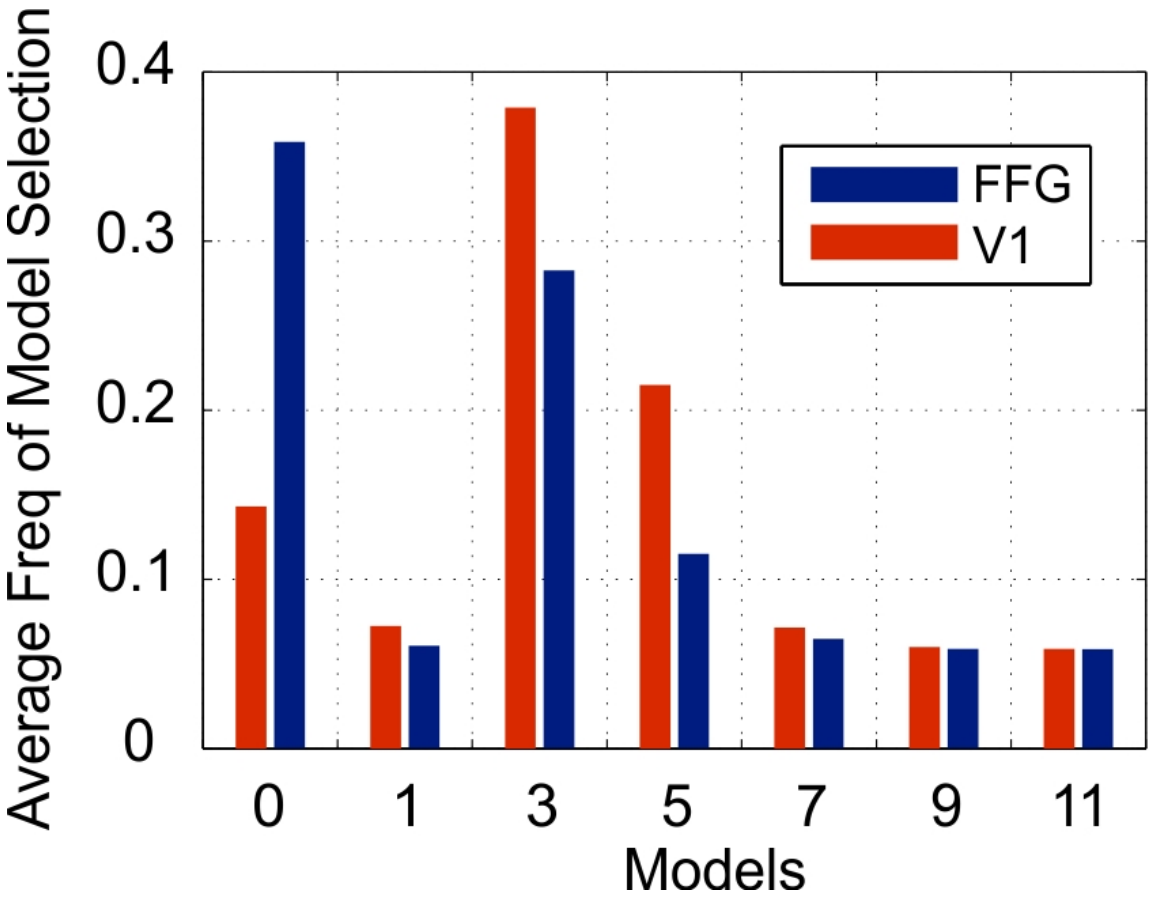

Figure 5.4: Average frequency of model selection based on population across each ROIs (FFG in blue and V1 in red). As demonstrated, model 3 (with feature dimension: normalized power in 3-10 Hz, 10-20 Hz, and 20-90 Hz frequency bands) is on average the most frequent selected model across the population in both FFG and V1.

Chapter 6

General discussion

6.1 Summary

The overarching aim of my PhD thesis was to develop a pipeline for decoding representations of visual stimuli, and to then use the pipeline to investigate temporal dynamics in the representation and replay of information, in episodic memory and working memory.

I first conducted an MEG experiment on episodic memory, in order to develop the pipeline tools (in chapter 2) and to investigate aspects of information replay (in chapter 3). The basis of the experiment was an associative recognition test, where, during an encoding phase, participants studied pairs of images and words (the images always being either faces or scenes, and the words always being semantically irrelevant to the images with which they were paired). Afterwards, participants were tested for word recognition and source memory, being given the associated word as a cue.

For the working memory MEG experiment, a paradigm that was a mixture of delay-

match-to-sample and Sternberg tests was designed, and used to study the oscillatory representations of three items from three visual categories (faces, fruits, and chairs). The following sections summarise the main findings of my thesis, and suggest possible implications and future directions for this research.

6.2 Decoding categorical representations using multivariate classifications

In line with some previous MEG decoding studies (e.g. Fuentemilla et al., 2010; Newman & Norman, 2010), for my thesis I selected, as features of interest for training classifiers, the power across (alpha/beta/gamma) frequency range during encoding (the time-period when the images were perceived) and across all the MEG channels. After testing multiple analytical approaches, chapter 2 showed that baseline-correcting the TF data does not improve, and possibly harms decoding performance. For example, I have shown that categorical-baseline correction (used for testing replay during working memory maintenance in Fuentemilla et al., 2010) is not appropriate for detecting the timing of representations of stimuli information during encoding, when using cross-validation. Instead, in this thesis I have suggested skipping a TF normalization step (see chapter 4), which is classically one of the preprocessing steps often used when studying induced responses.

Feature selection is an important step in pattern classification. In chapter 2 I demonstrated that categorical information for faces and scenes is evident in lower than gamma frequency (Gao et al., 2013), e.g. within a 8 to 45 Hz band, and I discussed the implications of this finding. Furthermore, I showed that including lower frequency ranges benefits decoding performance (in the same way that using lower

frequencies led to better classification performance in Carlson et al. (2013); van de Nieuwenhuijzen et al. (2013) studies).

There are a few caveats concerning the decoding pipeline in chapter 2, however. One limitation is based on the inconsistency across subjects in terms of which features were selected, and how many of these features were used at each time-point. Using t-tests as a feature selection step led to variable numbers of features across time-points. Such variation might affect classification performance. Although the SVM classification performance might minimally depend on the number of features (as shown for example in Besserve et al., 2007) a thorough investigation of this effect is not within the scope of this thesis.

Inconsistencies in feature selection, furthermore, have restricted the ability to form systematic conclusions on the impact of features on the representation of information. Addressing this issue, in chapter 5 I have proposed a group-level model selection approach for MEG multivariate analysis, which would enable discovery of a consistent feature space (across the population) with optimal decoding performance.

Another constraint of the decoding pipeline outlined in chapter 2 is time. Cross-validation is a time-consuming approach because feature selection, training and testing processes are repeated many times. In contrast, the method I have proposed in chapter 5 does not require cross-validation, and so it is n times (n = number of cross-validation folds) faster - I suggested using the Bayesian information criterion for assessing model fitness, and using Bayesian model selection for inferring accurate classification at the group-level.

In chapter 5 I used CVA for decoding. Unlike in chapter 2, spatial information

(channels or voxels) was not treated as a feature. Instead, the multivariate analysis was done independently on each source/voxel. In a future analysis, the multivariate analysis that I conducted in source space could be reformulated in a way to include spatial information as a feature; thus the characteristic oscillatory activity occurring across multiple brain regions for a particular class, condition, or stimulus could represent them.

6.3 Visual categorical information during encoding

In my thesis, consistent with previous findings (J. Liu et al., 2002; Gao et al., 2013), I have shown that multivariate decoding was able to extract the categorical information of stimuli at about 180 ms post-presentation, but not later, within a 1000 ms time-window during encoding (chapter 3 and chapter 4). Such early categorical information has been source-localized in IT and visual cortex (van de Nieuwenhuijzen et al., 2013). Additionally, in chapter 4 item-specific representations were decoded at 400 to 500 ms after the onset of the image. In order to understand the source of these representations, the features on which the SVM classifier relied would need to be extracted. One speculation I make is that there is a link between the timing of representations identified here and the well-known N400 ERP component, which has been suggested to represent semantic information (Ganis, Kutas, & Sereno, 1996), or to reflect recognition memory for familiar stimuli (in case of the frontal N400) (reviews: Rugg & Curran, 2007; Kutas & Federmeier, 2011).

6.4 Timing of pattern completion in episodic memory retrieval

Recollection is critically dependent on the hippocampal formation (Vargha-Khadem et al., 1997). In response to a partial memory-cue, this brain structure is capable of pattern-completion of the memory representations of the event (K. A. Norman, 2010; Kumaran & McClelland, 2012). In chapter 3, the partial memory cue I used was the associated word which was presented in isolation during retrieval phase. Here, it can be assumed that a pattern completion process was necessary, in order for retrieval of the image (that was paired with the word during encoding). To date, the timing of hippocampus-dependent pattern completion has remained unclear. That is because, thus far, it has not been possible to track the time course of reinstatements of single-event representations. A novel finding of my study in chapter 3 deals precisely with this issue, showing that such a reinstatement is quite rapid, occurring within a window of 446 to 513 ms after the onset of a partial memory cue.

Given that such an example of associative retrieval of scene information is likely to be hippocampus-dependent (Yonelinas et al., 2002; Horner et al., 2012), this finding indicates that hippocampus-dependent pattern completion processes, in response to partial memory cues, are more rapid than many previous ERP/ERF studies of recollection (e.g. Duzel, Vargha-Khadem, Heinze, & Mishkin, 2001; Addante, Ranganath, Olichney, & Yonelinas, 2012) have implied. In these studies, recollection was associated with ERP/ERF components emerging between 500 and 800 ms after presentation of the partial memory cue. Our data indicate that, within 500 ms, a hippocampal pattern completion process, and the ensuing cortical reinstatement, must have already reached the state of completion. Such rapid processing is consistent with more recent electromagnetic data from patients with bilateral hippocampal

lesions, showing that memory cues, such as used here, can initiate retrieval of contextual information within 350 ms (Horner et al., 2012). It is also consistent with recent intracortical recordings in humans showing that associative recognition effects appear at around 400ms in the perirhinal cortex, following earlier responses seen in the hippocampus at 250ms (Staresina et al., 2012).

MEG decoding was used in chapters 2 and 3 given that direct recordings from the human hippocampus are limited to clinical cases. In chapter 3 what I detected through decoding was most likely detected the replay of cortical representations. In order to study the process of hippocampal pattern completion, however, invasive recordings from the hippocampus are required, while subjects perform an associative recognition experiment as designed in chapter 3.

6.5 Recollection and familiarity

There is converging evidence suggesting that, during retrieval, some of the brain regions that were previously engaged during encoding are reactivated (Fenker, Schott, Richardson-Klavehn, Heinze, & Duzel, 2005; J. Johnson & Rugg, 2007; Donaldson, Wheeler, & Petersen, 2010; Chadwick, Hassabis, Weiskopf, & Maguire, 2010; Danker & Anderson, 2010). Consistent with this, animal studies have demonstrated that retrieving past events is associated with the reactivation of neural patterns elicited during study (Foster & Wilson, 2006; Carr, Jadhav, & Frank, 2011). Furthermore, an iEEG investigation of neural activity in the human hippocampus during free recall revealed that stimulus-specific hippocampal activity (occurring during presentation) was reinstated in a free recall task (e.g. Gelbard-Sagiv et al., 2008).

In fact, J. Johnson, Susan, Rugg, and Norman (2009) applied a Multi-Voxel Pattern Classifier on fMRI patterns of neural activity during an associative recognition task; interestingly, they showed that reinstatement of neural activity during retrieval was evident during both recollection and familiarity recognition. Accordingly, it was concluded that recollection and familiarity are continuous processes. However, this finding has not yet been replicated and it has remained unclear at what time point during retrieval neural activity patterns from encoding are reinstated.

In chapter 3 the number of source memory miss trials for word-scene associations was low (for 3 subjects it was fewer than 4 trials), and so this dataset did not allow me to systematically compare familiarity (word recognition without scene recall) and recollection (word recognition with scene recall). In my results for the word-face associations, I detected a significant correlation between the number of trials with high confidence recognition hits (remember) and the classification of words to their associated image category (faces) at 246 ms. This correlation could be related to a familiarity or recollection process, and this requires further systematic study. Here, it would be necessary to also account for the fact that faces were harder to remember than scenes.

Another emerging question with regards to the results of chapter 3 is whether, under circumstances where the retrieval of visual associations may not be experienced as recollection but is instead unconscious, reactivation could still be observed, and whether such reactivation would occur even earlier (i.e. at about 200 to 300 ms from onset of the stimuli (Waldhauser, Johansson, & Hanslmayr, 2012; Wimber et al., 2012)). Recent studies from the Hanslmeyer lab suggest this may be true (Waldhauser et al., 2012). However, our results from chapter 3 are not directly comparable to these studies because we used stimuli that were complex scenes and

faces, whereas Waldhauser and colleagues used basic visual features (e.g. colours). Future studies with a comparable number of trials with recollection and familiarity, and miss trials, should be used in a similar MEG decoding experiment, in order to examine the differences between replay that occurs during familiarity, and replay occurring during recollection.

6.6 Dependency of timing of recollection on saliency of the memory cue or saliency of the memory content

A life episode - for an episodic memory - is thought to be shaped by salient event boundaries (Ezzyat & Davachi, 2011); and some memory cues are more helpful for recollecting episodic memory than others (Ezzyat & Davachi, 2011). Ezzyat and Davachi (2011) have shown that the proportion of memory recall is higher when salient episodic boundaries are used as cues. An interesting question is whether the category-related differences that I found in chapter 3 are related to memory salience. In that study, scenes were better recollected than faces and the replay of scene information peaked at 513 ms, while for faces replay peaked earlier, at 446 ms; furthermore, the replay of faces at 513 ms was predictive of the overall performance of subjects in recollecting faces.

6.7 Timing of contextual representation

By definition, episodic memory includes information about *what*, *when* and *where* an event happened. The when- and where-ness are thought to shape the context of the event. In my thesis, I did not have the opportunity to study long-term memory for temporal information. However, I have investigated the neural mechanisms which may contribute to coding temporal information during perception and working memory maintenance. My rationale behind this approach was that, in order to represent context in episodic memory, it is necessary to keep track of the order of events within an episode. The findings in chapter 4 suggest that temporal context is shaped by the accumulation of information about the temporal order of a series of events, and that this accumulation is related to changes in working memory load.

In chapter 4 I found a main effect of the temporal order of stimuli presentations (or a main effect of working memory load) at about 300 to 500 ms after onset of the stimuli during encoding an effect akin to P300. In fact, there is converging evidence that the representations of working memory and of temporal context are closely linked. P300 is a well-known event-related response, reflecting working memory load and the characteristics of items kept in working memory (ERP; Morgan et al., 2008). For a long time, it has been suggested that the P300 response might reflect "context updating" (Donchin, 1981) and a closure of perceptual events (Verleger, 1988). Interestingly, Axmacher, Henseler, et al. (2010) have reported a P300-like stimulus related response in the hippocampus, which they related to increased theta power and phase coherence during early periods of maintenance. They also showed that hippocampal activity is reflective of working memory load during the gaps between presentations of successive stimuli (Axmacher et al., 2007). Hence, it is conceivable that working memory representations in the hippocampus relate to the temporal order (the context) of successive stimuli.

The working memory experiment in chapter 4 can contribute a further understanding of such mechanisms. It is possible that the theta activity which reflects temporal context (at 400 to 500 ms after the onset of the stimuli during encoding; chapter 4) is not only concurrent with the representations of stimuli (decoded at 400 to 500 ms) but is also involved in coordinating them. According to the theta-gamma coupling model of working memory maintenance, the hippocampus coordinates the order of sequential replay of information during the delay period (Lisman, 2010). The model predicts that during the maintenance, the information about sequential events is replayed in their original temporal order. Thus, a remaining question is whether stimulus representations are replayed according to their order of presentation.

To test this hypothesis, the decoders (or "marker") of stimulus specific information from the early (100 to 200 ms) and late (400 to 500 ms) time windows during encoding (in chapter 4) can be used for decoding when the memory load is more than 1. There are two predictions in here. Firstly, the emergence of representation of consequent items during encoding is phase-locked to ongoing theta oscillations (as shown in the representation of stimuli in the prefrontal cortex of monkey, Siegel et al., 2009); secondly, there is a sequential replay of working memory content during maintenance, which is phased-locked to theta oscillations. These predictions are along the lines of a temporal context model, whereby the temporal context is shaped by initial representations of the stimuli, which are time (or phase) dependent, and the replay of which provokes a (forward) sequential replay.

6.8 General conclusion

In my thesis I developed and used a method of decoding neural representations with high temporal resolution. I showed how categorical representations of stimuli (emerging at 180 ms after the onset of the stimuli during encoding) are conserved, and subsequently replayed during retrieval, leading to recollection of source memory. I showed that the timing of replay is from about 450 ms after the onset of a partial memory cue, which suggests that pattern completion (a process presumably dependent on hippocampal activity) is completed by that time. Furthermore, I also showed that in addition to categorically-specific representations of stimuli, item-specific representations (possibly semantic in nature) are detectable at about 400 ms after the onset of stimuli during encoding. These later representations of stimuli are concurrent with the timing of ERFs related to temporal order information or working memory load. Taken together, the MEG decoding methodology developed here provides a new avenue for future investigations of these memory mechanisms.

References

Addante, R. J., Ranganath, C., Olichney, J., & Yonelinas, A. P. (2012). Neurophysiological evidence for a recollection impairment in amnesia patients that leaves familiarity intact. *Neuropsychologia*. doi: 10.1016/j.neuropsychologia.2012.07.038

Aggleton, J., & Brown, M. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *The Behavioral and brain sciences*, *22*(3), 425-44; discussion 444. doi: 10.1017/s0140525x99002034

Amaral, D. (1999). Introduction: what is where in the medial temporal lobe? *Hippocampus*, *9*(1), 1-6. doi: 10.1002/(SICI)1098-1063(1999)9:1⟨1::AID-HIPO1⟩3.0.CO;2-T

Andersen, R., Snyder, L., Bradley, D., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual review of neuroscience*, *20*, 303-330. doi: 10.1146/annurev.neuro.20.1.303

Averbeck, B., Crowe, D., Chafee, M., & Georgopoulos, A. (2003). Neural activity in prefrontal cortex during copying geometrical shapes. II. decoding shape segments from neural ensembles. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, *150*(2), 142-153. doi: 10.1007/s00221-003-1417-5

Awh, E., Vogel, E., & Oh, S. (2006). Interactions between attention and working

memory. *Neuroscience*, *139*(1), 201-208. doi: 10.1016/j.neuroscience.2005.08.023

Axmacher, N., Henseler, M., Jensen, O., Weinreich, I., Elger, C., & Fell, J. (2010). Cross-frequency coupling supports multi-item working memory in the human hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(7), 3228-3233. doi: 10.1073/pnas.0911531107

Axmacher, N., Lenz, S., Haupt, S., Elger, C. E., & Fell, J. (2010). Electrophysiological signature of working and long-term memory interaction in the human hippocampus. *The European journal of neuroscience*, *31*(1), 177–188. doi: 10.1111/j.1460-9568.2009.07041.x

Axmacher, N., Mormann, F., Fernandez, G., Cohen, M., Elger, C., & Fell, J. (2007). Sustained neural activity patterns during working memory in the human medial temporal lobe. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *27*(29), 7807-7816. doi: 10.1523/JNEUROSCI.0962-07.2007

Baddeley, A. (1992). Working memory. *Science (New York, N.Y.)*, *255*(5044), 556-559. doi: 10.1126/science.1736359

Baddeley, A. (2010). Working memory. *Current biology : CB*, *20*(4), R136-R140. doi: 10.1016/j.cub.2009.12.014

Baeg, E., Kim, Y., Huh, K., Mook-Jung, I., Kim, H., & Jung, M. (2003). Dynamics of population code for working memory in the prefrontal cortex. *Neuron*, *40*(1), 177-188. doi: 10.1016/S0896-6273(03)00597-X

Barnes, G., Litvak, V., Brookes, M., & Friston, K. (2011). Controlling false positive rates in mass-multivariate tests for electromagnetic responses. *NeuroImage*, *56*(3), 1072-1081. doi: 10.1016/j.neuroimage.2011.02.072

Barnes, G. R., & Hillebrand, A. (2003). Statistical flattening of MEG beamformer images. *Human brain mapping*, *18*(1), 1–12. doi: 10.1002/hbm.10072

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working

memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10). doi: 10.1167/9.10.7

Berryhill, M., Phuong, L., Picasso, L., Cabeza, R., & Olson, I. (2007). Parietal lobe and episodic memory: bilateral damage causes impaired free recall of autobiographical memory. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *27*(52), 14415-14423. doi: 10.1523/JNEUROSCI .4163-07.2007

Besserve, M., Jerbi, K., Laurent, F., Baillet, S., Martinerie, J., & Garnero, L. (2007). Classification methods for ongoing EEG and MEG signals. *Biological research*, *40*(4), 415-437.

Bishop, C. (Ed.). (2006a). *Pattern recognition and machine learning*. Springer.

Bishop, C. (2006b). Pattern recognition and machine learning. In C. Bishop (Ed.), *Pattern recognition and machine learning* (chap. Continuous Latent Variables). Springer.

Bishop, C. (2006c). Pattern recognition and machine learning. In C. Bishop (Ed.), *Pattern recognition and machine learning* (p. 181-182). Springer.

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology*, *6*(2), 173-189.

Braver, T., Cohen, J., Nystrom, L., Jonides, J., Smith, E., & Noll, D. (1997). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage*, *5*(1), 49-62. doi: 10.1006/nimg.1996.0247

Brodersen, K. H., Mathys, C., Chumbley, J. R., Daunizeau, J., Ong, C. S., Buhmann, J. M., & Stephan, K. E. (2012). Bayesian mixed-effects inference on classification performance in hierarchical data sets. *Journal of Machine Learning Research*, *13*, 31333176.

Bruce, D. (2001). Fifty years since lashley's in search of the engram: refutations and conjectures. *Journal of the history of the neurosciences*, *10*(3), 308-318.

doi: 10.1076/jhin.10.3.308.9086

Buzsaki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science (New York, N.Y.)*, *304*(5679), 1926-1929. doi: 10.1126/science.1099745

Cabeza, R., Ciaramelli, E., Olson, I., & Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nature reviews. Neuroscience*, *9*(8), 613-625. doi: 10.1038/nrn2459

Cahill, L., & McGaugh, J. (1998). Mechanisms of emotional arousal and lasting declarative memory. *Trends in neurosciences*, *21*(7), 294-299.

Carl, C., Açk, A., Konig, P., Engel, A., & Hipp, J. (2012). The saccadic spike artifact in meg. *NeuroImage*, *59*(2), 1657-1667. doi: 10.1016/j.neuroimage .2011.09.020

Carlson, T., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal resolution decoding of object position and category. *Journal of vision*, *11*(10). doi: 10.1167/11.10.9

Carlson, T., Tovar, D., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: the first 1000 ms. *Journal of vision*, *13*(10). doi: 10.1167/13.10.1

Carr, M., Jadhav, S., & Frank, L. (2011). Hippocampal replay in the awake state: a potential substrate for memory consolidation and retrieval. *Nature neuroscience*, *14*(2), 147-153. doi: 10.1038/nn.2732

Cashdollar, N., Malecki, U., Fergus, R., Duncan, J., Lavie, N., & Duzel, E. (2009). Hippocampus-dependent and -independent theta-networks of active maintenance. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(48), 20493-20498. doi: 10.1073/pnas.0904823106

Chadwick, M., Hassabis, D., Weiskopf, N., & Maguire, E. (2010). Decoding individual episodic memory traces in the human hippocampus. *Current biology : CB*, *20*(6), 544-547. doi: 10.1016/j.cub.2010.01.053

Chalise, P., & Fridley, B. (2012). Comparison of penalty functions for sparse canonical correlation analysis. *Computational Statistics and Data Analysis*, *56*, 245?254.

Chatfield, C., & Collins, A. J. (1980). *Introduction to multivariate analysis*. Chapman and Hall, London.

Christopher, M., Lepage, K., Eden, U., & Eichenbaum, H. (2011). Hippocampal "time cells" bridge the gap in memory for discontiguous events. *Neuron*, *71*(4), 737-749. doi: 10.1016/j.neuron.2011.07.012

Ciaramelli, E., Grady, C., & Moscovitch, M. (2008). Top-down and bottom-up attention to memory: a hypothesis (AtoM) on the role of the posterior parietal cortex in memory retrieval. *Neuropsychologia*, *46*(7), 1828-1851. doi: 10.1016/j.neuropsychologia.2008.03.022

Cohen, J., Perlstein, W., Braver, T., Nystrom, L., Noll, D., Jonides, J., & Smith, E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, *386*(6625), 604-608. doi: 10.1038/386604a0

Colby, C. (1998). Action-oriented spatial reference frames in cortex. *Neuron*, *20*(1), 15-24.

Colby, C., & Goldberg, M. (1999). Space and attention in parietal cortex. *Annual review of neuroscience*, *22*, 319-349. doi: 10.1146/annurev.neuro.22.1.319

Corkin, S., Amaral, D., Gonzalez, R., Johnson, K., & Hyman, B. (1997). H. m.'s medial temporal lobe lesion: findings from magnetic resonance imaging. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *17*(10), 3964-3979.

Craik, F. (2002). Levels of processing: past, present. and future? *Memory (Hove, England)*, *10*(5-6), 305-318. doi: 10.1080/09658210244000135

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, *11*(6), 671-684.

Curran, T., Tanaka, J., & Weiskopf, D. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cognitive, affective & behavioral neuroscience*, *2*(1), 1-18.

Dalal, S., Baillet, S., Adam, C., Ducorps, A., Schwartz, D., Jerbi, K., ... Lachaux, J. (2009). Simultaneous MEG and intracranial EEG recordings during attentive reading. *NeuroImage*, *45*(4), 1289-1304. doi: 10.1016/j.neuroimage.2009.01.017

Danker, J., & Anderson, J. (2010). The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. *Psychological bulletin*, *136*(1), 87-8102. doi: 10.1037/a0017937

Davachi, L., Mitchell, J., & Wagner, A. (2003). Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(4), 2157-2162. doi: 10.1073/pnas.0337195100

Davachi, L., & Wagner, A. (2002). Hippocampal contributions to episodic encoding: insights from relational and item-based learning. *Journal of neurophysiology*, *88*(2), 982-990.

Deffke, I., Sander, T., Heidenreich, J., Sommer, W., Curio, G., Trahms, L., & Lueschow, A. (2007). MEG/EEG sources of the 170-ms response to faces are co-localized in the fusiform gyrus. *NeuroImage*, *35*(4), 1495-1501. doi: 10.1016/j.neuroimage.2007.01.034

Dien, J., Michelson, C., & Franklin, M. (2010). Separating the visual sentence n400 effect from the p400 sequential expectancy effect: cognitive and neuroanatomical implications. *Brain research*, *1355*, 126-140. doi: 10.1016/j.brainres.2010.07.099

Dolan, R., Lane, R., Chua, P., & Fletcher, P. (2000). Dissociable temporal lobe activations during emotional episodic memory retrieval. *Neuroimage*, *11*(3),

203-209.

Donaldson, D., Petersen, S., Ollinger, J., & Buckner, R. (2001). Dissociating state and item components of recognition memory using fMRI. *NeuroImage*, *13*(1), 129-142. doi: 10.1006/nimg.2000.0664

Donaldson, D., Wheeler, M., & Petersen, S. (2010). Remember the source: dissociating frontal and parietal contributions to episodic memory. *Journal of cognitive neuroscience*, *22*(2), 377-391. doi: 10.1162/jocn.2009.21242

Donchin, E. (1981). Presidential address, 1980. Surprise!...Surprise? *Psychophysiology*, *18*(5), 493-513. doi: 10.1111/j.1469-8986.1981.tb01815.x

Downing, P. (2000). Interactions between visual working memory and selective attention. *Psychological science*, *11*(6), 467-473. doi: 10.1111/1467-9280.00290

Duarte, A., Ranganath, C., & Knight, R. (2005). Effects of unilateral prefrontal lesions on familiarity, recollection, and source memory. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *25*(36), 8333-8337. doi: 10.1523/JNEUROSCI.1392-05.2005

Dudai, Y. (2012). The restless engram: consolidations never end. *Annual review of neuroscience*, *35*, 227–247. doi: 10.1146/annurev-neuro-062111-150500

Duncan, K. K., Hadjipapas, A., Li, S., Kourtzi, Z., Bagshaw, A., & Barnes, G. (2010). Identifying spatially overlapping local cortical networks with MEG. *Human Brain Mapping*, *31*(7), 1003–1016. doi: 10.1002/hbm.20912

Duzel, E., Habib, R., Schott, B., Schoenfeld, A., Lobaugh, N., McIntosh, A., ... Heinze, H. (2003). A multivariate, spatiotemporal analysis of electromagnetic time-frequency data of recognition memory. *NeuroImage*, *18*(2), 185-197.

Duzel, E., Neufang, M., & Heinze, H.-J. (2005). The oscillatory dynamics of recognition memory and its relationship to event-related responses. *Cerebral cortex (New York, N.Y. : 1991)*, *15*(12), 1992-2002. doi: 10.1093/cercor/bhi074

Duzel, E., Penny, W., & Burgess, N. (2010). Brain oscillations and memory. *Current*

opinion in neurobiology, *20*(2), 143-149. doi: 10.1016/j.conb.2010.01.004

Duzel, E., Picton, T., Cabeza, R., Yonelinas, A., Scheich, H., Heinze, H., & Tulving, E. (2001). Comparative electrophysiological and hemodynamic measures of neural activation during memory-retrieval. *Human brain mapping*, *13*(2), 104-123. doi: 10.1002/hbm.1028

Duzel, E., Vargha-Khadem, F., Heinze, H., & Mishkin, M. (2001). Brain activity evidence for recognition without recollection after early hippocampal damage. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(14), 8101-8106. doi: 10.1073/pnas.131205798

Egorov, A., Hamam, B., Fransen, E., Hasselmo, M., & Alonso, A. (2002). Graded persistent activity in entorhinal cortex neurons. *Nature*, *420*(6912), 173-178. doi: 10.1038/nature01171

Eichenbaum, H. (2008). *Learning & memory*. New York: W.W. Norton & Co.

Everling, S., Tinsley, C., Gaffan, D., & Duncan, J. (2006). Selective representation of task-relevant objects and locations in the monkey prefrontal cortex. *The European journal of neuroscience*, *23*(8), 2197-2214. doi: 10.1111/j.1460-9568.2006.04736.x

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological science*, *22*(2), 243-252. doi: 10.1177/0956797610393742

Fell, J., & Axmacher, N. (2011). The role of phase synchronization in memory processes. *Nature Reviews Neuroscience*, *12*(2), 105–118. doi: 10.1038/nrn2979

Fell, J., Klaver, P., Elfadil, H., Schaller, C., Elger, C., & Fernández, G. (2003). Rhinal-hippocampal theta coherence during declarative memory formation: interaction with gamma synchronization? *The European journal of neuroscience*, *17*(5), 1082-1088. doi: 10.1046/j.1460-9568.2003.02522.x

Fell, J., Klaver, P., Lehnertz, K., Grunwald, T., Schaller, C., Elger, C., & Fernandez, G. (2001). Human memory formation is accompanied by rhinal-hippocampal

coupling and decoupling. *Nature neuroscience*, *4*(12), 1259-1264. doi: 10.1038/nn759

Fenker, D., Schott, B., Richardson-Klavehn, A., Heinze, H.-J., & Duzel, E. (2005). Recapitulating emotional context: activity of amygdala, hippocampus and fusiform cortex during recollection and familiarity. *The European journal of neuroscience*, *21*(7), 1993-1999. doi: 10.1111/j.1460-9568.2005.04033.x

Fernndez, G., Effern, A., Grunwald, T., Pezer, N., Lehnertz, K., Dmpelmann, M., ... Elger, C. E. (1999). Real-time tracking of memory formation in the human rhinal cortex and hippocampus. *Science (New York, N.Y.)*, *285*(5433), 1582–1585.

Fisch, L., Privman, E., Ramot, M., Harel, M., Nir, Y., Kipervasser, S., ... Malach, R. (2009). Neural "ignition": enhanced activation linked to perceptual awareness in human ventral stream visual cortex. *Neuron*, *64*(4), 562-574. doi: 10.1016/j.neuron.2009.11.001

Foster, D., & Wilson, M. (2006). Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature*, *440*(7084), 680-683. doi: 10.1038/nature04587

Freedman, D., Riesenhuber, M., Poggio, T., & Miller, E. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science (New York, N.Y.)*, *291*(5502), 312-316. doi: 10.1126/science.291.5502.312

Fries, P., Fernndez, G., & Jensen, O. (2003). When neurons form memories. *Trends in neurosciences*, *26*(3), 123–124. doi: 10.1016/S0166-2236(03)00023-7

Friston, K. J., Ashburner, J., Kiebel, S. J., Nichols, T. E., & Penny, W. D. (Eds.). (2007). *Statistical parametric mapping: The analysis of functional brain images*. Academic Press. Software and data available from http://www.fil.ion.ucl.ac.uk/spm/.

Fuentemilla, L., Penny, W., Cashdollar, N., Bunzeck, N., & Duzel, E. (2010). Theta-

coupled periodic replay in working memory. *Current biology : CB*, *20*(7), 606-612. doi: 10.1016/j.cub.2010.01.057

Ganis, G., Kutas, M., & Sereno, M. (1996). The search for common sense: an electrophysiological study of the comprehension of words and pictures in reading. *Journal of cognitive neuroscience*, *8*(2), 89-8106. doi: 10.1162/jocn.1996.8.2.89

Gao, Z., Goldstein, A., Harpaz, Y., Hansel, M., Elana, Z., & Bentin, S. (2013). A magnetoencephalographic study of face processing: M170, gamma-band oscillations and source localization. *Human brain mapping*, *34*(8), 1783-1795. doi: 10.1002/hbm.22028

Gelbard-Sagiv, H., Mukamel, R., Harel, M., Malach, R., & Fried, I. (2008). Internally generated reactivation of single neurons in human hippocampus during free recall. *Science (New York, N.Y.)*, *322*(5898), 96–101. doi: 10.1126/science.1164685

Gevins, A., Smith, M., Leong, H., L, M., Whitfield, S., Du, R., & Rush, G. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human factors*, *40*(1), 79-91. doi: 10.1518/001872098779480578

Graf, P., & Schacter, D. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of experimental psychology. Learning, memory, and cognition*, *11*(3), 501-518.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual review of neuroscience*, *27*, 649-677. doi: 10.1146/annurev.neuro.27.070203.144220

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 694-699. doi: 10.1073/pnas.98.2.694

Hannula, D., & Ranganath, C. (2008). Medial temporal lobe activity predicts successful relational memory binding. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *28*(1), 116-124. doi: 10.1523/JNEUROSCI.3086-07.2008

Hanslmayr, S., Spitzer, B., & Bauml, K. (2009). Brain oscillations dissociate between semantic and nonsemantic encoding of episodic memories. *Cerebral cortex (New York, N.Y. : 1991)*, *19*(7), 1631-1640. doi: 10.1093/cercor/bhn197

Hanslmayr, S., Staudigl, T., & Fellner, M.-C. (2012). Oscillatory power decreases and long-term memory: the information via desynchronization hypothesis. *Frontiers in human neuroscience*, *6*, 74. doi: 10.3389/fnhum.2012.00074

Harrison, S., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632-635. doi: 10.1038/nature07832

Hartley, T., Bird, C., Chan, D., Cipolotti, L., Husain, M., Faraneh, V., & Burgess, N. (2007). The hippocampus is required for short-term topographical memory in humans. *Hippocampus*, *17*(1), 34-48. doi: 10.1002/hipo.20240

Hasselmo, M., & Stern, C. (2006). Mechanisms underlying working memory for novel information. *Trends in cognitive sciences*, *10*(11), 487-493. doi: 10.1016/j.tics.2006.09.005

Haxby, J., Grady, C., Horwitz, B., Ungerleider, L., Mishkin, M., Carson, R., ... Rapoport, S. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(5), 1621-1625. doi: 10.1073/pnas.88.5.1621

Haynes, J., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. (2007). Reading hidden intentions in the human brain. *Current biology : CB*, *17*(4), 323-328. doi: 10.1016/j.cub.2006.11.072

Henson, R., Rugg, M., Shallice, T., & Dolan, R. (2000). Confidence in recognition memory for words: dissociating right prefrontal roles in episodic retrieval. *Journal of cognitive neuroscience*, *12*(6), 913-923. doi: 10.1162/08989290051137468

Hoffman, K. L., & McNaughton, B. L. (2002). Coordinated reactivation of distributed memory traces in primate neocortex. *Science (New York, N.Y.)*, *297*(5589), 2070–2073. doi: 10.1126/science.1073538

Hoogenboom, N., Schoffelen, J.-M., Oostenveld, R., Parkes, L., & Fries, P. (2006). Localizing human visual gamma-band activity in frequency, time and space. *NeuroImage*, *29*(3), 764-773. doi: 10.1016/j.neuroimage.2005.08.043

Horner, A. J., Gadian, D. G., Fuentemilla, L., Jentschke, S., Vargha-Khadem, F., & Duzel, E. (2012). A rapid, hippocampus-dependent, item-memory signal that initiates context memory in humans. *Current biology: CB*, *22*(24), 2369–2374. doi: 10.1016/j.cub.2012.10.055

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*, 269–299. doi: 10.1006/jmps.2001.1388

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, *96*(16), 9379–9384. doi: 10.1073/pnas.96.16.9379

Jafarpour, A., Horner, A. J., Fuentemilla, L., Penny, W. D., & Duzel, E. (2013). Decoding oscillatory representations and mechanisms in memory. *Neuropsychologia*, *51*(4), 772–780. doi: 10.1016/j.neuropsychologia.2012.04.002

Jenkins, L., & Ranganath, C. (2010). Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *30*(46), 15558-15565. doi:

10.1523/JNEUROSCI.1337-10.2010

Jensen, O., & Colgin, L. (2007). Cross-frequency coupling between neuronal oscillations. *Trends in cognitive sciences*, *11*(7), 267-269. doi: 10.1016/j.tics.2007.05.003

Jensen, O., Gelfand, J., Kounios, J., & Lisman, J. (2002). Oscillations in the alpha band (9-12 hz) increase with memory load during retention in a short-term memory task. *Cerebral cortex (New York, N.Y. : 1991)*, *12*(8), 877-882. doi: 10.1093/cercor/12.8.877

Jensen, O., & Tesche, C. (2002). Frontal theta activity in humans increases with memory load in a working memory task. *The European journal of neuroscience*, *15*(8), 1395-1399. doi: 10.1046/j.1460-9568.2002.01975.x

Johnson, A., van der Meer, M. A. A., & Redish, A. D. (2007). Integrating hippocampus and striatum in decision-making. *Curr Opin Neurobiol*, *17*(6), 692697. doi: 10.1016/j.conb.2008.01.003

Johnson, J., & Rugg, M. (2007). Recollection and the reinstatement of encoding-related cortical activity. *Cerebral cortex (New York, N.Y. : 1991)*, *17*(11), 2507-2515. doi: 10.1093/cercor/bhl156

Johnson, J., Susan, M., Rugg, M., & Norman, K. (2009). Recollection, familiarity, and cortical reinstatement: a multivoxel pattern analysis. *Neuron*, *63*(5), 697-708. doi: 10.1016/j.neuron.2009.08.011

Jones, M., & Wilson, M. (2005). Theta rhythms coordinate hippocampal-prefrontal interactions in a spatial memory task. *PLoS biology*, *3*(12). doi: 10.1371/journal.pbio.0030402

Kahn, I., Davachi, L., & Wagner, A. D. (2004). Functional-neuroanatomic correlates of recollection: implications for models of recognition memory. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *24*(17), 4172–4180. doi: 10.1523/JNEUROSCI.0624-04.2004

Kandel, E., Schwartz, J., & Jessell, T. (2000). *Principals of neural science*. McGraw-Hill.

Kanwisher, N., Stanley, D., & Harris, A. (1999). The fusiform face area is selective for faces not animals. *Neuroreport*, *10*(1), 183-187. doi: 10.1097/00001756-199901180-00035

Kilner, J. M., Kiebel, S. J., & Friston, K. J. (2005). Applications of random field theory to electrophysiology. *Neuroscience letters*, *374*(3), 174–178. doi: 10.1016/j.neulet.2004.10.052

Klimesch, W., Doppelmayr, M., Schwaiger, J., Winkler, T., & Gruber, W. (2000). Theta oscillations and the ERP old/new effect: independent phenomena? *Clinical Neurophysiology*, *111*(5), 781-793.

Klimesch, W., Schimke, H., & Schwaiger, J. (1994). Episodic and semantic memory: an analysis in the EEG theta and alpha band. *Electroencephalography and clinical neurophysiology*, *91*(6), 428-441.

Knowlton, B., & Squire, L. (1995). Remembering and knowing: two different expressions of declarative memory. *Journal of experimental psychology. Learning, memory, and cognition*, *21*(3), 699-710. doi: 10.1037/0278-7393.21.3.699

Kravitz, D., Kriegeskorte, N., & Baker, C. (2010). High-level visual object representations are constrained by position. *Cerebral cortex (New York, N.Y. : 1991)*, *20*(12), 2916-2925. doi: 10.1093/cercor/bhq042

Kreiman, G., Koch, C., & Fried, I. (2000a). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature neuroscience*, *3*(9), 946-953. doi: 10.1038/78868

Kreiman, G., Koch, C., & Fried, I. (2000b). Imagery neurons in the human brain. *Nature*, *408*(6810), 357-361. doi: 10.1038/35042575

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United*

States of America, *103*(10), 3863–3868. doi: 10.1073/pnas.0600244103

Kriegeskorte, N., Mur, M., Ruff, D., Kiani, R., Bodurka, J., Esteky, H., ... Bandettini, P. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126-1141. doi: 10.1016/j.neuron.2008.10.043

Kucera, H., & Francis, W. (1967). *Computational analysis of present-day american english*. Brown Univ. Press, Providence, RI.

Kuhl, B., Rissman, J., & Wagner, A. (2012). Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory. *Neuropsychologia*, *50*(4), 458-469. doi: 10.1016/j.neuropsychologia.2011.09.002

Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5903–5908. doi: 10.1073/pnas.1016939108

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. *Psychological review*, *119*(3), 573–616. doi: 10.1037/a0028681

Kutas, M., & Federmeier, K. (2011). Thirty years and counting: finding meaning in the n400 component of the event-related brain potential (ERP). *Annual review of psychology*, *62*, 621-647. doi: 10.1146/annurev.psych.093008.131123

Lashley, K. S. (1950). In search of the engram. *Symp. Soc. Exp. Biol.*, *4*, 454–482.

Lee, A., & Wilson, M. (2002). Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron*, *36*(6), 1183-1194.

Lee, H., Simpson, G., Logothetis, N., & Rainer, G. (2005). Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex. *Neuron*, *45*(1), 147-156. doi: 10.1016/j.neuron.2004.12

.025

Lega, B. C., Jacobs, J., & Kahana, M. (2012). Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus*, *22*(4), 748–761. doi: 10.1002/hipo.20937

Leshikar, E., & Duarte, A. (2012). Medial prefrontal cortex supports source memory accuracy for self-referenced items. *Social neuroscience*, *7*(2), 126-145. doi: 10.1080/17470919.2011.585242

Lisman, J. (2010). Working memory: the importance of theta and gamma oscillations. *Current biology*, *20*(11), R490-R492. doi: 10.1016/j.cub.2010.04.011

Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., ... Friston, K. (2011). EEG and MEG data analysis in SPM8. *Computational intelligence and neuroscience*, *2011*, 852961. doi: 10.1155/2011/852961

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: an MEG study. *Nature neuroscience*, *5*(9), 910-916. doi: 10.1038/nn909

Liu, X., Ramirez, S., Pang, P. T., Puryear, C. B., Govindarajan, A., Deisseroth, K., & Tonegawa, S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature*, *484*(7394), 381–385. doi: 10.1038/nature11028

Llera, A., van Gerven, M., Gómez, V., Jensen, O., & Kappen, H. (2011). On the use of interaction error potentials for adaptive brain computer interfaces. *Neural networks : the official journal of the International Neural Network Society*, *24*(10), 1120-1127. doi: 10.1016/j.neunet.2011.05.006

Lopes da Silva, F. (2006). Event-related neural activities: what about phase? *Progress in brain research*, *159*, 3-17. doi: 10.1016/S0079-6123(06)59001-6

Manning, J., Polyn, S., Baltuch, G., Litt, B., & Kahana, M. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National Academy of Sciences of the United States of Amer-*

ica, *108*(31), 12893-12897. doi: 10.1073/pnas.1015174108

Marr, D. (1971). Simple memory: a theory for archicortex. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *262*(841), 23-81.

McCarthy, G., Puce, A., Belger, A., & Allison, T. (1999). Electrophysiological studies of human face perception. II: response properties of face-specific potentials generated in occipitotemporal cortex. *Cerebral cortex (New York, N.Y. : 1991)*, *9*(5), 431-444. doi: 10.1093/cercor/9.5.431

McClelland, J., B, M., & OReilly, R. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, *102*(3), 419-457. doi: 10.1037/0033-295x.102.3.419

McEvoy, L., Smith, M., & Gevins, A. (1998). Dynamic cortical networks of verbal and spatial working memory: effects of memory load and task practice. *Cerebral cortex (New York, N.Y. : 1991)*, *8*(7), 563-574. doi: 10.1093/cercor/8.7.563

Miller, E., & Cohen, J. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, *24*, 167-202. doi: 10.1146/annurev.neuro.24.1.167

Mishkin, M. (1982). A memory system in the monkey. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *298*(1089), 83-95. doi: 10.1098/rstb.1982.0074

Mizuhara, H., & Yamaguchi, Y. (2011). Neuronal ensemble for visual working memory via interplay of slow and fast oscillations. *The European journal of neuroscience*, *33*(10), 1925-1934. doi: 10.1111/j.1460-9568.2011.07681.x

Morgan, H., Klein, C., Boehm, S., Shapiro, K., & Linden, D. (2008). Working memory load for faces modulates p300, n170, and n250r. *Journal of cognitive neuroscience*, *20*(6), 989-981002. doi: 10.1162/jocn.2008.20072

Mormann, F., Fell, J., Axmacher, N., Weber, B., Lehnertz, K., Elger, C., & Fernan-

dez, G. (2005). Phase/amplitude reset and theta-gamma interaction in the human medial temporal lobe during a continuous word recognition memory task. *Hippocampus*, *15*(7), 890-900. doi: 10.1002/hipo.20117

Morris, R. (2006). Elements of a neurobiological theory of hippocampal function: the role of synaptic plasticity, synaptic tagging and schemas. *The European journal of neuroscience*, *23*(11), 2829-2846. doi: 10.1111/j.1460-9568.2006.04888.x

Morris, R., Inglis, J., Ainge, J., Olverman, H., Tulloch, J., Dudai, Y., & Kelly, P. (2006). Memory reconsolidation: sensitivity of spatial memory to inhibition of protein synthesis in dorsal hippocampus during encoding and retrieval. *Neuron*, *50*(3), 479-489. doi: 10.1016/j.neuron.2006.04.012

Moscovitch, M., Rosenbaum, R., Gilboa, A., Addis, D., Westmacott, R., Grady, C., ... Nadel, L. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. *Journal of anatomy*, *207*(1), 35-66. doi: 10.1111/j.1469-7580.2005.00421.x

Mosher, J., Leahy, R., & Lewis, P. (1999). Eeg and meg forward solutions for inverse methods. *IEEE transactions on bio-medical engineering*, *46*(3), 245-259.

Murakami, S., & Okada, Y. (2006). Contributions of principal neocortical neurons to magnetoencephalography and electroencephalography signals. *The Journal of physiology*, *575*(Pt 3), 925-936. doi: 10.1113/jphysiol.2006.105379

Nadel, L., & Hardt, O. (2011). Update on memory systems and processes. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, *36*(1), 251-273. doi: 10.1038/npp.2010.169

Nadel, L., Hupbach, A., Gomez, R., & K, N. (2012). Memory formation, consolidation and transformation. *Neuroscience and biobehavioral reviews*, *36*(7), 1640-1645. doi: 10.1016/j.neubiorev.2012.03.001

Neufang, M., Heinze, H., & Duzel, E. (2006). Electromagnetic correlates of recognition memory processes. *Clinical EEG and neuroscience : official journal*

of the EEG and Clinical Neuroscience Society (ENCS), *37*(4), 300-308. doi: 10.1177/155005940603700407

Newman, E., & Norman, K. (2010). Moderate excitation leads to weakening of perceptual representations. *Cerebral cortex (New York, N.Y. : 1991)*, *20*(11), 2760-2770. doi: 10.1093/cercor/bhq021

Nieder, A., Freedman, D., & Miller, E. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science (New York, N.Y.)*, *297*(5587), 1708-1711. doi: 10.1126/science.1072493

Nolte, G., & Curio, G. (2000). Current multipole expansion to estimate lateral extent of neuronal activity: a theoretical analysis. *IEEE transactions on bio-medical engineering*, *47*(10), 1347-1355. doi: 10.1109/10.871408

Norman, K., Detre, G., & Polyn, S. (2008). The cambridge handbook of computational psychology. In R. Sun (Ed.), *The cambridge handbook of computational psychology.* (chap. Computational models of episodic memory). Cambridge: Cambridge University Press.

Norman, K. A. (2010). How hippocampus and cortex contribute to recognition memory: revisiting the complementary learning systems model. *Hippocampus*, *20*(11), 1217–1227. doi: 10.1002/hipo.20855

Nunez, P. (1981). *Electric fields of the brain.* Oxford, New York: Oxford University Press.

Nystrom, L., Braver, T., Sabb, F., Delgado, M., Noll, D., & Cohen, J. (2000). Working memory for letters, shapes, and locations: fMRI evidence against stimulus-based regional organization in human prefrontal cortex. *NeuroImage*, *11*(5 Pt 1), 424-446. doi: 10.1006/nimg.2000.0572

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain research*, *34*(1), 171-175.

Olson, I., Moore, K., Stark, M., & Chatterjee, A. (2006). Visual working memory is impaired when the medial temporal lobe is damaged. *Journal of cognitive neuroscience*, *18*(7), 1087-1097. doi: 10.1162/jocn.2006.18.7.1087

Onton, J., & Makeig, S. (2006). Information-based modeling of event-related brain dynamics. *Progress in brain research*, *159*, 99–120. doi: 10.1016/S0079-6123(06)59007-7

OReilly, R., & Rudy, J. (2000). Computational principles of learning in the neocortex and hippocampus. *Hippocampus*, *10*(4), 389-397. doi: 10.1002/1098-1063(2000)10:4⟨389::AID-HIPO5⟩3.0.CO;2-P

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Comput.*, *18*(2), 283328. doi: 10.1162/089976606775093909

Osipova, D., Takashima, A., Oostenveld, R., Fernandez, G., Maris, E., & Jensen, O. (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *26*(28), 7523-7531. doi: 10.1523/JNEUROSCI.1948-06.2006

Peelen, M., & Downing, P. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of neurophysiology*, *93*(1), 603-608. doi: 10.1152/jn.00513.2004

Penny, W., & Flandin, G. (2005). Bayesian analysis of fMRI data with spatial priors. In *Proceedings of the joint statistical meeting (JSM).* American Statistical Association.

Penny, W. D., & Roberts, S. J. (1998). Bayesian neural networks for classification: how useful is the evidence framework ? *Neural Networks*, *12*, 877–892.

Poch, C., Fuentemilla, L., Barnes, G., & Duzel, E. (2011). Hippocampal theta-phase modulation of replay correlates with configural-relational short-term memory performance. *The Journal of neuroscience : the official journal of the Society*

for Neuroscience, *31*(19), 7038-7042. doi: 10.1523/JNEUROSCI.6305-10.2011

Polyn, S., Natu, V., Cohen, J., & Norman, K. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science (New York, N.Y.)*, *310*(5756), 1963-1966. doi: 10.1126/science.1117645

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological review*, *116*(1). doi: 10.1037/a0014420

Press, W., Flannery, B., Teukolsky, S., & Vetterling, W. (1991). *Numerical recipes in c*. Cambridge University Press.

Puce, A., Allison, T., & McCarthy, G. (1999). Electrophysiological studies of human face perception. III: effects of top-down processing on face-specific potentials. *Cerebral cortex (New York, N.Y.: 1991)*, *9*(5), 445–458.

Quiroga, R., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*(7045), 1102-1107. doi: 10.1038/nature03687

Rainer, G., Asaad, W., & Miller, E. (1998). Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature*, *393*(6685), 577-579. doi: 10.1038/31235

Reynolds, J., & O'Reilly, R. (2009). Developing pfc representations using reinforcement learning. *Cognition*, *113*(3), 281-292. doi: 10.1016/j.cognition.2009.05.015

Ritchey, M., Wing, E. A., Labar, K. S., & Cabeza, R. (2012). Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. *Cerebral cortex (New York, N.Y.: 1991)*. doi: 10.1093/cercor/bhs258

Robinson, S., & Vrba, J. (1999). Recent advances in biomagnetism. In T. Yoshimoto, M. Kotani, S. Kuriki, H. Karibe, & N. Nakasato (Eds.), *Recent advances in biomagnetism* (p. 302305). Sendai, Japan: Tohoku Univ. Press.

Rosa, M. J., Bestmann, S., Harrison, L., & Penny, W. (2010). Bayesian model selection maps for group studies. *Neuroimage*, *49*(1), 217224. doi: 10.1016/j.neuroimage.2009.08.051

Rossion, B., & Caharel, S. (2011). ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision research*, *51*(12), 1297–1311. doi: 10.1016/j.visres.2011.04.003

Rugg, M., Cox, C., Doyle, M., & Wells, T. (1995). Event-related potentials and the recollection of low and high frequency words. *Neuropsychologia*, *33*(4), 471-484.

Rugg, M., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in cognitive sciences*, *11*(6), 251-257. doi: 10.1016/j.tics.2007.04.004

Rugg, M., & Yonelinas, A. (2003). Human recognition memory: a cognitive neuroscience perspective. *Trends in cognitive sciences*, *7*(7), 313-319.

Rypma, B., & M, D. (1999). The roles of prefrontal brain regions in components of working memory: effects of memory load and individual differences. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(11), 6558-6563. doi: 10.1073/pnas.96.11.6558

Saksida, L., & Bussey, T. (2010). The representational-hierarchical view of amnesia: translation from animal to human. *Neuropsychologia*, *48*(8), 2370-2384. doi: 10.1016/j.neuropsychologia.2010.02.026

Schacter, D. L., & Graf, P. (1986). Effects of elaborative processing on implicit and explicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(3), 432–444. doi: 10.1037/0278-7393.12.3.432

Schindler, A., & Bartels, A. (2013). Parietal cortex codes for egocentric space beyond the field of view. *Current biology : CB*, *23*(2), 177-182. doi: 10.1016/

j.cub.2012.11.060

Scoville, W., & Milner, B. (2000). Loss of recent memory after bilateral hippocampal lesions. 1957. *The Journal of neuropsychiatry and clinical neurosciences*, *12*(1), 103-113.

Sederberg, P., Kahana, M., Howard, M., Donner, E., & Madsen, J. (2003). Theta and gamma oscillations during encoding predict subsequent recall. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *23*(34), 10809-10814. doi: 10.1038/nrn1303

Sederberg, P., Schulze-Bonhage, A., Madsen, J., Bromfield, E., McCarthy, D., Brandt, A., ... Kahana, M. (2007). Hippocampal and neocortical gamma oscillations predict memory formation in humans. *Cerebral cortex (New York, N.Y. : 1991)*, *17*(5), 1190-1196. doi: 10.1093/cercor/bhl030

Sekihara, K., Nagarajan, S. S., Poeppel, D., Marantz, A., & Miyashita, Y. (2002). Application of an MEG eigenspace beamformer to reconstructing spatio-temporal activities of neural sources. *Human brain mapping*, *15*(4), 199–215.

Sereno, A., & Lehky, S. (2011). Population coding of visual space: comparison of spatial representations in dorsal and ventral pathways. *Frontiers in computational neuroscience*, *4*. doi: 10.3389/fncom.2010.00159

Sereno, M., Dale, A., Reppas, J., Kwong, K., Belliveau, J., Brady, T., ... Tootell, R. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science (New York, N.Y.)*, *268*(5212), 889-893. doi: 10.1126/science.7754376

Siegel, M., Warden, M., & Miller, E. (2009). Phase-dependent neuronal coding of objects in short-term memory. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(50), 21341-21346. doi: 10.1073/pnas.0908193106

Sirota, A., Montgomery, S., Fujisawa, S., Isomura, Y., Zugaro, M., & Buzsaki, G.

(2008). Entrainment of neocortical neurons and gamma oscillations by the hippocampal theta rhythm. *Neuron*, *60*(4), 683-697. doi: 10.1016/j.neuron .2008.09.014

Snyder, L. (2000). Coordinate transformations for eye and arm movements in the brain. *Current opinion in neurobiology*, *10*(6), 747-754.

Sohn, M.-H., Goode, A., Stenger, V., Carter, C., & Anderson, J. (2003). Competition and representation during memory retrieval: roles of the prefrontal cortex and the posterior parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(12), 7412-7417. doi: 10.1073/ pnas.0832374100

Soto, J., Pantazis, D., Jerbi, K., Lachaux, J., Garnero, L., & Leahy, R. (2009). Detection of event-related modulations of oscillatory brain activity with multivariate statistical analysis of MEG data. *Human brain mapping*, *30*(6), 1922-1934. doi: 10.1002/hbm.20765

Sperling, R., Chua, E., Cocchiarella, A., Erin, R., Poldrack, R., Schacter, D., & Albert, M. (2003). Putting names to faces: successful encoding of associative memories activates the anterior hippocampal formation. *NeuroImage*, *20*(2), 1400-1410. doi: 10.1016/S1053-8119(03)00391-4

Squire, L., Stark, C., & Clark, R. (2004). The medial temporal lobe. *Annual review of neuroscience*, *27*, 279-306. doi: 10.1146/annurev.neuro.27.070203.144130

Squire, L., Wixted, J., & Clark, R. (2007). Recognition memory and the medial temporal lobe: a new perspective. *Nature reviews. Neuroscience*, *8*(11), 872-883. doi: 10.1038/nrn2154

Staresina, B. P., Henson, R. N. A., Kriegeskorte, N., & Alink, A. (2012). Episodic reinstatement in the medial temporal lobe. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, *32*(50), 18150–18156. doi: 10.1523/JNEUROSCI.4156-12.2012

Stephan, K., Penny, W., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *Neuroimage*, *46*(4), 1004–17.

Stern, C., Sherman, S., Kirchhoff, B., & Hasselmo, M. (2001). Medial temporal and prefrontal contributions to working memory tasks with novel and familiar stimuli. *Hippocampus*, *11*(4), 337-346. doi: 10.1002/hipo.1048

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*(3736), 652–654. (PMID: 5939936) doi: 10.1126/science.153.3736.652

Sutherland, G., & McNaughton, B. (2000). Memory trace reactivation in hippocampal and neocortical neuronal ensembles. *Current opinion in neurobiology*, *10*(2), 180-186.

Tallon-Baudry, & Bertrand. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in cognitive sciences*, *3*(4), 151-162.

Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Permier, J. (1997). Oscillatory gamma-band (30-70 hz) activity induced by a visual search task in humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *17*(2), 722-734.

Tallon-Baudry, C., Bertrand, O., Peronnet, F., & Pernier, J. (1998). Induced gamma-band activity during the delay of a visual short-term memory task in humans. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *18*(11), 4244-4254.

Tipping, M. E., & Faul, A. C. (2003). Proceedings of the ninth international workshop on artificial intelligence and statistics. In C. Bishop & B. Frey (Eds.), *Proceedings of the ninth international workshop on artificial intelligence and statistics* (chap. Fast marginal likelihood maximisation for sparse Bayesian models). FL: Key West.

Treves, A., & Rolls, E. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*(3), 374-391. doi: 10.1002/hipo.450040319

Tubridy, S., & Davachi, L. (2011). Medial temporal lobe contributions to episodic sequence encoding. *Cerebral cortex (New York, N.Y. : 1991)*, *21*(2), 272-280. doi: 10.1093/cercor/bhq092

Tuladhar, A., ter Huurne, N., Schoffelen, J., Maris, E., Oostenveld, R., & Jensen, O. (2007). Parieto-occipital sources account for the increase in alpha activity with working memory load. *Human brain mapping*, *28*(8), 785-792. doi: 10.1002/hbm.20306

Tulving, E. (1985). How many memory systems are there? *American Psychologist*, *40*(4), 385-398. doi: 10.1037/0003-066X.40.4.385

Tulving, E., & Donaldson, W. (1972). *Organization of memory.* Academic Press.

Ungerleider, L. G., & Haxby, J. V. (1994). What and where in the human brain. *Current Opinion in Neurobiology*, *4*, 157–165. doi: 10.1016/0959-4388(94)90066-3

van de Nieuwenhuijzen, M. E., Backus, A. R., Bahramisharif, A., Doeller, C. F., Jensen, O., & van Gerven, M. A. J. (2013). MEG-based decoding of the spatiotemporal dynamics of visual category perception. *NeuroImage*. doi: 10.1016/j.neuroimage.2013.07.075

Van Essen, D. C., & Maunsell, J. H. (1983). Hierarchical organization and functional streams in the visual cortex. *Trends in neurosciences*, *6*, 370-375.

VanRullen, R., & Thorpe, S. (2001). The time course of visual processing: from early perception to decision-making. *Journal of cognitive neuroscience*, *13*(4), 454-461. doi: 10.1162/08989290152001880

Vapnik, V. (2000a). The nature of statistical learning theory. In (p. 138-140). springer.

Vapnik, V. (2000b). *The nature of statistical learning theory.* springer.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential effects of early hippocampal pathology

on episodic and semantic memory. *Science (New York, N.Y.)*, *277*(5324), 376–380.

Verleger, R. (1988). Event-related potentials and cognition: A critique of the context updating hypothesis and an alternative interpretation of p3. *Behavioral and brain sciences*, *11*(3), 343-356.

Vertes, R. (2005). Hippocampal theta rhythm: a tag for short-term memory. *Hippocampus*, *15*(7), 923-935. doi: 10.1002/hipo.20118

Vogel, E., & Machizawa, M. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*(6984), 748-751. doi: 10.1038/nature02447

Voss, J., & Paller, K. (2009). Remembering and knowing: electrophysiological distinctions at encoding but not retrieval. *NeuroImage*, *46*(1), 280-289. doi: 10.1016/j.neuroimage.2009.01.048

Wagner, A. D. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, *281*. doi: 10.1126/science.281.5380.1188

Waldhauser, G., Johansson, M., & Hanslmayr, S. (2012). / oscillations indicate inhibition of interfering visual memories. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *32*(6), 1953-1961. doi: 10.1523/JNEUROSCI.4201-11.2012

Wang, C. (2007). Variational bayesian approach to canonical correlation analysis. *IEEE Trans. Neur. Networks*, *18*(3), 905910. doi: 10.1109/TNN.2007.891186

Wang, S., & Morris, R. (2010). Hippocampal-neocortical interactions in memory formation, consolidation, and reconsolidation. *Annual review of psychology*, *61*, 49-79, C1. doi: 10.1146/annurev.psych.093008.100523

Wang, W., Sudre, G., Xu, Y., Kass, R., Collinger, J., Degenhart, A., ... Weber, D. (2010). Decoding and cortical source localization for intended movement

direction with MEG. *Journal of neurophysiology*, *104*(5), 2451-2461. doi: 10.1152/jn.00239.2010

Waydo, S., Kraskov, A., Quian Quiroga, R., Fried, I., & Koch, C. (2006). Sparse representation in the human medial temporal lobe. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *26*(40), 10232-10234. doi: 10.1523/JNEUROSCI.2101-06.2006

Weiss, S., Muller, H., & Rappelsberger, P. (2000). Theta synchronization predicts efficient memory encoding of concrete and abstract nouns. *Neuroreport*, *11*(11), 2357-2361. doi: 10.1097/00001756-200008030-00005

Wimber, M., Maaß, A., Staudigl, T., Alan, R., & Hanslmayr, S. (2012). Rapid memory reactivation revealed by oscillatory entrainment. *Current biology : CB*, *22*(16), 1482-1486. doi: 10.1016/j.cub.2012.05.054

Woodruff, C., Hayama, H., & Rugg, M. (2006). Electrophysiological dissociation of the neural correlates of recollection and familiarity. *Brain research*, *1100*(1), 125-135. doi: 10.1016/j.brainres.2006.05.019

Yassa, M., & Stark, C. (2011). Pattern separation in the hippocampus. *Trends in neurosciences*, *34*(10), 515-525. doi: 10.1016/j.tins.2011.06.006

Yonelinas, A., Kroll, N., Quamme, J., Lazzara, M., Sauvé, M., Widaman, K., & Knight, R. (2002). Effects of extensive temporal lobe damage or mild hypoxia on recollection and familiarity. *Nature neuroscience*, *5*(11), 1236-1241. doi: 10.1038/nn961

Yonelinas, A., Otten, L., Shaw, K., & Rugg, M. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *25*(11), 3002-3008. doi: 10.1523/JNEUROSCI.5295-04.2005

Yonelinas, A., & Parks, C. (2007). Receiver operating characteristics (ROCs) in recognition memory: a review. *Psychological bulletin*, *133*(5), 800-832. doi:

10.1037/0033-2909.133.5.800

Zhou, J. (2008). Dimension reduction based on constrained canonical correlation and variable filtering. *The Annals of Statistics*, *36*(4), 1649–1668. (Mathematical Reviews number (MathSciNet): MR2435451; Zentralblatt MATH identifier: 1142.62045) doi: 10.1214/07-AOS529